

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 09:14:42 ; Search time 2079 Seconds

(without alignments)
5229.234 Million cell updates/sec

Title: US-09-714-865b-1

Perfect score: 1 acttgaagtcacatg999g.....aagtcgtggtttgatgca 2224

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3216467 seqs, 244149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Listing first 45 summaries

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US04_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US02_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US01_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US04_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US02_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US01_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1601	72.0	1984	16	US-10-104-047-1652 Sequence 1652, Ap
2	389.8	17.5	420	9	US-09-833-381-1475 Sequence 1475, Ap
3	386.4	17.4	2451	14	US-10-044-090-480 Sequence 480, App
4	384.8	17.3	4423	14	US-10-002-600-67 Sequence 67, Appl
5	383.2	17.2	4416	15	US-10-293-582-4 Sequence 4, Appl1
6	356	16.0	591	16	US-09-764-891-1170 Sequence 1170, Ap
7	327	13.7	2019	15	US-10-032-585-6564 Sequence 6564, Ap
8	301.6	13.6	2362	17	US-10-437-963-5848 Sequence 5848, Ap
9	298.2	13.4	1815	16	US-10-369-493-46077 Sequence 46077, A
10	298	13.4	1854	16	US-10-369-493-46167 Sequence 46167, A
11	294.6	13.2	1908	16	US-10-369-493-46378 Sequence 46378, A
12	291.4	13.1	3012	13	US-10-424-599-14656 Sequence 14656, A
13	282.2	12.7	2672	13	US-10-424-599-14660 Sequence 14660, A
14	276.2	12.4	1681	16	US-10-369-493-26954 Sequence 26954, A

15	275.8	12.4	1902	9	US-09-938-842A-521 Sequence 521, App
16	275.8	12.4	1902	11	US-09-938-842A-521 Sequence 521, App
17	273	12.3	539	15	US-10-029-386-12207 Sequence 12207, A
18	272.8	12.3	1705	16	US-10-260-338-490 Sequence 490, App
19	271	12.2	271	15	US-10-029-386-25907 Sequence 25907, A
20	268.4	12.1	2025	15	US-10-128-714-2551 Sequence 2551, Ap
21	268.4	12.1	2025	15	US-10-128-714-2551 Sequence 2551, Ap
22	268.4	12.1	2105	15	US-10-128-714-1551 Sequence 1551, Ap
23	268.4	12.1	2105	15	US-10-128-714-6551 Sequence 6551, Ap
24	268.4	12.1	4105	15	US-10-128-714-5551 Sequence 5551, Ap
25	268.4	12.1	4105	15	US-10-128-714-5551 Sequence 5551, Ap
26	263.2	11.8	1519	13	US-10-424-599-9376 Sequence 9376, Ap
27	229.8	10.3	2519	17	US-10-437-963-60284 Sequence 60284, A
28	220	9.9	2021	16	US-10-320-797-1250 Sequence 1250, Ap
29	220	9.9	2194	10	US-09-764-891-7481 Sequence 7481, Ap
30	220	9.9	2195	10	US-09-764-891-7482 Sequence 7482, Ap
31	220	9.9	2197	10	US-09-764-891-7483 Sequence 7483, Ap
32	220	9.9	4021	16	US-10-320-797-250 Sequence 250, App
33	219.2	9.9	1860	16	US-10-320-797-2250 Sequence 2250, Ap
34	210.2	9.5	3841	15	US-10-198-846-12716 Sequence 12716, A
35	206	9.3	2619	17	US-10-437-963-46500 Sequence 46500, A
36	204.8	9.2	1884	17	US-10-437-963-19215 Sequence 19215, A
37	191.4	8.6	1606	13	US-10-424-599-21465 Sequence 21465, A
38	188.8	8.5	2596	13	US-10-425-114-22685 Sequence 22685, A
39	178.2	8.0	2236	13	US-10-424-599-55854 Sequence 55854, A
40	175.2	7.9	1962	13	US-10-425-114-31433 Sequence 31433, A
41	173	7.8	2934	17	US-10-437-963-69859 Sequence 69859, A
42	172	7.7	1538	13	US-10-425-114-13560 Sequence 13560, A
43	172	7.7	1504	13	US-10-425-114-23650 Sequence 23650, A
44	170.2	7.7	1176	16	US-10-369-493-28786 Sequence 28786, A
45	168.4	7.6	403	9	US-09-960-352-14361 Sequence 14361, A

ALIGNMENTS

RESULT 1
US-10-104-047-1652
; Sequence 1652, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1652
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1652

Query Match 72.0%; Score 1601; DB 16; Length 1984;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1662; Conservative 0; Mismatches 5; Indels 48; Gaps 1;

QY	510	ATATGACTTAAACCCAGAGCAATGATGACGACGCTGTGCTTTTGTCTTAA	569	Sequence 521, App
				Sequence 12207, A
DB	117	ATATGACTTAAACCCAGAGCAATGATGACGACGCTGTGCTTTTGTCTTAA	176	Sequence 490, App
				Sequence 25907, A
QY	570	GACCAATTAATGACGACGATGATGATCTTCTTAAAGCAGAGTGACGTGAA	629	Sequence 2551, Ap
				Sequence 7551, Ap
DB	177	GACCAATTAATGACGACGATGATGATGATCTTCTTAAAGCAGAGTGACGTGAA	236	Sequence 1551, Ap
				Sequence 6551, Ap
QY	630	GTCAGAGGTGCTTCAAAAGTTTAAATGAAGATTAATACAGGCTCGAAATTT	689	Sequence 551, App
				Sequence 551, Ap
DB	237	GTCAGAGGTGCTTCAAAAGTTTAAATGAAGATTAATACAGGCTCGAAATTT	248	Sequence 9376, Ap
				Sequence 60284, A
QY	690	CTTGAAGTCAGAGCAGAGGAGGAGAAATGATGATCTCAAGACCAAAAGTACCT	749	Sequence 1250, Ap
				Sequence 7482, Ap
				Sequence 7483, Ap
				Sequence 250, App
				Sequence 2250, Ap
				Sequence 12716, A
				Sequence 46500, A
				Sequence 19215, A
				Sequence 21465, A
				Sequence 22685, A
				Sequence 55854, A
				Sequence 31433, A
				Sequence 69859, A
				Sequence 13560, A
				Sequence 23650, A
				Sequence 28786, A
				Sequence 14361, A

```

Db      249  |||||CTTGAAGTCAGAAAGCAGAAAGAGAAAGTGTGATCTCGAGAGCAAAAGTGACCT 308
Qy      750  |||||ACATACCCCTCCCTCCACCTGAGAGATGAGACCTCATCTTTGGCATTATACAGACGGCA 809
Db      309  |||||ACATACCCCTCCCTCCACCTGAGAGATGAGACCTCATCTTTGGCATTATACAGACGGCA 368
Qy      810  |||||TAAACTTCGACAAATAGACACTATCTTTGAGAAAGTGTCTGACATGATGACACACAG 869
Db      369  |||||TAACTTCGACAAATAGACACTATCTTTGAGAAAGTGTCTGACATGATGACACACAG 428
Qy      870  |||||CAATTCGACTTTGAGAAAGCTATCTCTGTGACAGACATGATTAACAATTTGCTTAAG 929
Db      429  |||||CAATTCGACTTTGAGAAAGCTATCTCTGTGACAGACATGATTAACAATTTGCTTAAG 488
Qy      930  |||||CTGGTTACTAAGCTTACTCTGTGCAAAAATACAGTATCTCATCACTTGCAGAGAC 989
Db      489  |||||CTGGTTACTAAGCTTACTCTGTGCAAAAATACAGTATCTCATCACTTGCAGAGAC 548
Qy      990  |||||GAGATTGATGCTTTGCTCAAAACAGGCTCTGGAAAGACTGGGCTTTTCTCTTACCA 1049
Db      549  |||||GAGATTGATGCTTTGCTCAAAACAGGCTCTGGAAAGACTGGGCTTTTCTCTTACCA 608
Qy      1050 |||||TTTGGGCTCATATGATGATGATGAAATTAATGCACTGCTTTTAAAGATTGACAGAAC 1109
Db      609  |||||TTTGGGCTCATATGATGATGATGAAATTAATGCACTGCTTTTAAAGATTGACAGAAC 668
Qy      1110 |||||CAGAGTGTATTATTGAGCAACAACTCGAGAAATTTGTCAACAGATTATTGGAAGCA 1169
Db      669  |||||CAGAGTGTATTATTGAGCAACAACTCGAGAAATTTGTCAACAGATTATTGGAAGCA 728
Qy      1170 |||||GAAATTTTCTTTGGGACTTTGTGTAGAGCTGTGTATATATATGAGGGGAAACCCAGCTGG 1229
Db      729  |||||GAAATTTTCTTTGGGACTTTGTGTAGAGCTGTGTATATATATGAGGGGAAACCCAGCTGG 788
Qy      1230 |||||GACATTCATTCGACAAATAGTACAGAGCTGTATATATATATGAGCTGTCTGAGAGAC 1289
Db      789  |||||GACATTCATTCGACAAATAGTACAGAGCTGTATATATATATGAGCTGTCTGAGAGAC 848
Qy      1290 |||||TGAATGATATCATATAGGCAAAAGAAATTTGCTCTCAACAGATCAATATCTTATGTTTGG 1349
Db      849  |||||TGAATGATATCATATAGGCAAAAGAAATTTGCTCTCAACAGATCAATATCTTATGTTTGG 908
Qy      1350 |||||ATGAGCTGATCCAGATTTGATGATGAGTTTGTGTCTCAAGAAATGAGAAATTAATTTCTT 1409
Db      909  |||||ATGAGCTGATCCAGATTTGATGATGAGTTTGTGTCTCAAGAAATGAGAAATTAATTTCTT 968
Qy      1410 |||||GCCAGAGATTCATCAAGAGAAACAGGCCAAACCTTATGTTCACTGCAACTTTTCCAG 1469
Db      969  |||||GCCAGAGATTCATCAAGAGAAACAGGCCAAACCTTATGTTCACTGCAACTTTTCCAG 1028
Qy      1470 |||||AGGAAATTCAAAGTTGGCTGACAGAGTTTAAAGTCAATTAATCTGTTTGTGCTGTTG 1529
Db      1029 |||||AGGAAATTCAAAGTTGGCTGACAGAGTTTAAAGTCAATTAATCTGTTTGTGCTGTTG 1088
Qy      1530 |||||GACAAGTGGGTGAGCATGTAGAGATGTTCAAGACCCGTTCTCCAAAGTTGGCCAGTTCT 1589
Db      1089 |||||GACAAGTGGGTGAGCATGTAGAGATGTTCAAGACCCGTTCTCCAAAGTTGGCCAGTTCT 1148
Qy      1590 |||||CAAAAAGAGAAAGCTCGTTGAAATCTCGGAAACATATAGGGAATGAATAATATAGTCT 1649
Db      1149 |||||CAAAAAGAGAAAGCTCGTTGAAATCTCGGAAACATATAGGGAATGAATAATATAGTCT 1208
Qy      1650 |||||TTGTGAACTAAGAAAAAGAGATTTTACTGCAACTTTTCTTGTCAAGAAAAATAT 1709
Db      1209 |||||TTGTGAACTAAGAAAAAGAGATTTTACTGCAACTTTTCTTGTCAAGAAAAATAT 1268
Qy      1710 |||||CAACTACAGTATCCATGTGATTCGGGAAACAGAGACGCGGAGCAAGCTCTTTGGAGATT 1769
Db      1269 |||||CAACTACAGTATTCATGTGATTCGGGAAACAGAGACGCGGAGCAAGCTCTTTGGAGATT 1328
Qy      1770 |||||TTGGCTTTGAAAGTCCCAAGTTCTTGTGCTACTTCACTGACCTGCAAGAGGCTGAGATA 1829

```

```

Db      1329 |||||TTGCTTTGAAAGTGCACCAAGTCTTGTGTCTACTTCACTAGTACTGTCAGAGGCTGATTA 1388
Qy      1830 |||||TTGAAATAGTGCACATATGTTATCAATTTTGAATCTTCCCTTACACATTTGATTAATGTTTC 1889
Db      1389 |||||TTGAAATAGTGCACATATGTTATCAATTTTGAATCTTCCCTTACACATTTGATTAATGTTTC 1448
Qy      1890 |||||ATCGAATTTGGGCGTACTGCTGTGTGTGGAAATCTGCGACAGCAATTTCTTTTGTATC 1949
Db      1449 |||||ATCGAATTTGGGCGTACTGCTGTGTGTGGAAATCTGCGACAGCAATTTCTTTTGTATC 1508
Qy      1950 |||||TTGAATCGGATTAACATTTAGCACAGCTCTTAGTAAAGTATTGACAGATGCTCAACAGG 2009
Db      1509 |||||TTGAATCGGATTAACATTTAGCACAGCTCTTAGTAAAGTATTGACAGATGCTCAACAGG 1568
Qy      2010 |||||ATGTTCTGCACTGTTTGGAAAGAAATTTGCTTTTACTACATACATTTCTGCTTCACTGCTGA 2069
Db      1569 |||||ATGTTCTGCACTGTTTGGAAAGAAATTTGCTTTTACTACATACATTTCTGCTTCACTGCTGA 1628
Qy      2070 |||||GTACAAAGGAAACGCTGTTTGCATCAGTTGATACCAAGAAAGGCAAGACCTTTGAACA 2129
Db      1629 |||||GTACAAAGGAAACGCTGTTTGCATCAGTTGATACCAAGAAAGGCAAGACCTTTGAACA 1688
Qy      2130 |||||CAGCTGGGTTTCTTCTTCAAGAGCTCCCAATCCAGTAGATGATGATGATTTAA 2189
Db      1689 |||||CAGCTGGGTTTCTTCTTCAAGAGCTCCCAATCCAGTAGATGATGATGATTTAA 1748
Qy      2190 |||||GCCAAAACATCTTCAAGCTGTGTGTTTGAATGCA 2224
Db      1749 |||||GCCAAAACATCTTCAAGCTGTGTGTTTGAATGCA 1783

RESULT 2
US-09-833-381-1475
; Sequence 1475, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2000-02-29
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1475
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1475

Query Match 17.5%; Score 389.8; DB 9; Length 420;
Best Local Similarity 99.5%; Pred. No. 3.1e-100;
Matches 391; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      818  |||||GACAAATACGACACTATCTTTGTGGAAGTGTGACATGATGACACCAAGCAATTTCTG 877
Db      28   |||||GTCAATATACGACACTATCTTTGTGGAAGTGTGACATGATGACACCAAGCAATTTCTG 87
Qy      878  |||||ACTTTTGAAGAAGTATCTCTGTGACAGACTGAATTAACAATTTGCTTAAAGCTGTTAT 937
Db      88   |||||ACTTTTGAAGAAGTATCTCTGTGACAGACTGAATTAACAATTTGCTTAAAGCTGTTAT 147
Qy      938  |||||ACTAAGCTTACTCTGTGCAAAAATACAGATTTCTATCATACTTGTGACAGAGAGATTGG 997
Db      148  |||||ACTAAGCTTACTCTGTGCAAAAATACAGATTTCTATCATACTTGTGACAGAGAGATTGG 207
Qy      998  |||||ATGCTTTGTGCTCAACAGGCTCTGGAGACATGCGGCTTTTCTCTCAATTTTGGCT 1057
Db      208  |||||ATGCTTTGTGCTCAACAGGCTCTGGAGACATGCGGCTTTTCTCTCAATTTTGGCT 267
Qy      1058 |||||CATATGATGATGATGGAATTAATGCGCAGTGTATTAAGATTTGACAGAAACAGAGTGT 1117

```



```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 350667.1
US-10-002-600-67

```

Query Match	17.3%;	Score 384.8;	DB 14;	Length 4423;
Best Local Similarity	58.6%;	Pred. No. 4.1e-98;		
Matches 762;	Conservative	0;	Mismatches 502;	Indels 36;
				Gaps 4;

QY	80	AGCAGAGCAATTAACCTCGACAAATACGACATCTATCTTGGAAAGTGTGACATGAG	860
Dp	536	ACACGGGGATTAACTTTGAGAAATATGATGATATACAGATGAGGACACCGGAGTA	595
QY	861	CACCAACGACAAATTCGACTTTTGAAGAGCTAATCTGTCAACATGATATACACA	920
Dp	596	GTCTCCACATATTGAGAAATTTTACCGATATTGACATGGAGAAATATATCTGCGACA	655
QY	921	TTGCTMAACCTGATATACATAAGCTTACTCTGTGCAAAAATPACGATATCCATGAC	980
Dp	656	TTGAACTTACTGGCTTACTGCTGTCTACTCCAGTGCAGAAAACATGCCATTTCTATTAT	715
QY	981	TTGCAGACGAGATTGATGCGCTGTGCTCAACAGGCTGGGAGAACTCGGCTTTTC	1040
Dp	716	AGGAAAAAGACATTTAAATGCTGTGCGCCCAACAGATCTGGGAAAACTGCAGCATTC	775
QY	1041	TCCTACCAATTTTGGCTCA-----TATGATCAATGAG	1072
Dp	776	TTTTTACCCACTAGTACGATATATACATGCTCCAGAGAACTTTGAAAGCTGTGA	835
QY	1074	GAATAACTGCCAGTCGTTTAAAGATTGCAGAAACCAAGTATATTATTGACCCA	1133
Dp	836	AGGAATATGAAAGGATATGGGCGCCGCAAAACATATCCATATCTTGTTTAGCCCAA	895
QY	1134	CTCGAATATTGCTCAACCGAATTTATTTGAAAGCCAAAAATTTCTTTTGGACCTGTG	1192
Dp	896	CAAGAAATATGCTGTACAGATCTATGAGAGGCAAAAAATTTCTTACCGATCTGAG	955
QY	1194	TAAAGCTGTGTTATATATATGAGGGAACCCAGCTGGACATTCATTTGCACAAATAGTAC	1253
Dp	956	TTCTGCTCTGTGTAGTTTATGATGCTGATATTTGGTCACAGATTCCGGACTTAAAC	1015
QY	1254	AAGGCTGTATATATTATATGCTACTCTCTGGAAACATGATGATATCATAGCCAAAGAA	1313
Dp	1016	GTGGATGCCACTGTGATATACCCACTCCAGACGCTCTAGTGATATGATGAAAGAGAA	1075
QY	1314	AGATTGATCTCAACAGATCAAAACTAGTTTGGATGAAAGCTGATCCAGTTGAGATA	1374
Dp	1076	AGATTGATATACCTTCGCAAGTACTTAGTGTGATGAAAGCTGATGAGATGCTGATTA	1133
QY	1374	TGGCTTTTGGTCCAGAAATGAAGAATTATTTCTTSCCGAGAAATGCCATCAAGAAAC	1433
Dp	1136	TGGGATTTGAACTTCAGATACGTCGTATGTTGMAACAAGTACTATGACCACAAAAGGGG	1192
QY	1434	AGGCGCAAAACCTTATGTTTCAGTGCACATTTCCAGAGAAATTCAAAGTTGGCTGAC	1493
Dp	1196	TTCTGTCACACCATGATGTTTAGTGTACTTTTCTTAAGAAATACAGATGCTTGTCTG	1255
QY	1494	AGTTTAAAGTCAAATTATCTGTTGTTGCTGTGGCAAGTGGGTGAGCATGTAGAG	1553
Dp	1256	ACTTTTGGATGAA---TATATCTTTTGGCTGTGAGCAGATGAGGCTCTACTCTCGAGA	1312
QY	1554	ATGTTCCAGAGACGGTCTCCAAATGGCCAGTTCTCAAAABAGAAAGCTGTGAAA	1613
Dp	1313	ACATCACACAGAAATAGTTGGGTGGAAGCTTAGATTAACGGTCAATTTCTACTGAGCA	1372
QY	1614	TTCTCGAAACATAGGG---GATGAAAGAACTATGCTTTTGTGGAACTAAGAAAAAG	1670
Dp	1373	TTTTTAGTGCACACAGGAGTATTCATCTACTTAGTGTGGAGACCAAAAAGGGAG	1432
QY	1671	CAGATTATACGAACTTTCTTTGTCAAGAAAAAATATCAATCAATATCATGTGT	1733
Dp	1433	CAGATTCCCTGAGGATTTCTTAACATGAAAGATATGTTGTACTAGTATTCATGAG	1492

Qy	1731	ATCGGAAACAGAGAGACCGGAGCAACTCTTGAGAGATTTCGTTGGAAAGGCCAG	1790
Db	1493	ACCGCTACAGAGAGATCGAGAGGAGGCCCTTACCAAGTTTCGCTCAGAGAAAAGCCCA	1552
Qy	1791	TTCTGTTGCTACTCTTCAGTAGCTGCCAGAGGCGCTGGAATTGGAAAATGTCAATGTGA	1850
Db	1553	TTCTAGTGGCTAACAGCTGTGGCAGCACAGAGACCTAGACATTTCAAATGTAGAGATGTGA	1612
Qy	1851	TCATATTTGATCTCTCTTCCATACACATTGATATATGTTCATCGAATTGGGGGTACTGTC	1910
Db	1613	TCATATTTGATTTGCCAAGTGATATTTAGAAATATGTGACATCGATTGGCCGTACAGAC	1672
Qy	1911	GTTCGTGGAACTACGGAGAGCAATTTCCCTTTTGTACTTGAATCTTGAATCGGATTAACATTTAG	1970
Db	1673	GTTGAGAAACCTGGGCGCTTGCCACTCTCATTTCTTTA--TGAATAAAATTTGAAATTTTA	1729
Qy	1971	CACAGCCTCTAGTAAAAGTATTTGACAGATGCTCAACAGATGTTCCTGCATGGTTGGAAG	2030
Db	1730	CAAAGGATTTGTGAGATCTCTCTGTAGAACTTAACAGAGAGTGCCCTTCTGTGGTGAAA	1789
Qy	2031	AAATTTGCTTTAGACATACATTCCTGGCTTTCAGTGTAG	2070
Db	1790	ATATGCTTATGAACACCACTACAGAGGTGGCACTGCTGG	1829

```

RESULT 5
US-10-293-582-4
Sequence 4, Application US/10293582
Publication No. US20030175253A1
GENERAL INFORMATION:
APPLICANT: Akil, Huda
APPLICANT: Bunney, William E.
APPLICANT: Burke, Sharon
APPLICANT: Choudary, Prabhakara V.
APPLICANT: Cox, David R.
APPLICANT: Evans, Simon
APPLICANT: Jones, Edward G.
APPLICANT: Li, Tun
APPLICANT: Lopez, Juan F.
TITLE OF INVENTION: The Trustees of The Leland Stanford Junior University
TITLE OF INVENTION: Compositions and Methods for Diagnosing and Treating
TITLE OF INVENTION: Mental Disorders
FILE REFERENCE: 020685-000210US
CURRENT APPLICATION NUMBER: US/10/293,582
CURRENT FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/339,252
PRIOR FILING DATE: 2001-11-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 4416
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: dead box, y isoform (DBY) RNA helicas
US-10-293-582-4

```

Query Match	17.2%	Score 393.2	DB 15	Length 4416
Best Local Similarity	58.5%	Prod. 1.2e-97		
Matches 761	Conservative 0	Mismatches 503	Indels 36	Gaps 4

QY	801	AGACAGGCATPAACTTCGACCAATACGACCTATTTCTGTGGAAGTGTCTGACATGATG	860
Db	529	ACACGGGGATTACTTTGAGAAATATGATATACGATGAGGCAACCGGCAATACT	588
QY	861	CACCAACCAAGCAATTCGACTTTTGAGGAAGCTATCTCTGTGAGACCTGAAATPACACA	920
Db	589	GTCCTCACATATTGGAAATTTTGGCATTTATGACATGGAGAGAAATTAATCATGCGGACACA	648
QY	921	TTGCTAAAGTGGTTTACTPAAAGCTACTCCGTGCAAAAATACAGATATTCATCATAC	980
Db	649	TTTAACCTTACTGCTACTACTCGTCTACTCCAAATGCAAAAACATGCCATTCCTATTATTA	708

QY 981 TTGCGAGACAGATTTGATGGCTTGTGCTCAACAGAGGTCGGGAGACTGCGGCTTTTC 1040
 Db 709 AGGAAAAAAGAGCTTAGTGGCTTGTGCCAAACAGAGATCGGAAAACTCGACGATTTTC 768
 QY 1041 TCCACCAATTTTGGCTCA-----TATGATGATGATG 1073
 Db 769 TTTTACCACTAGAGTCAGATATATATACAGATGTCAGAGAGAGAGCTTTGAAGCTGTGA 828
 QY 1074 GAATACCTGCACTGCTTTTAAAGAGTTCAGAGAACCAAGTGTATATTTGACGCCAA 1133
 Db 829 AGGAAATGAGAGTATGCGCGCGCAACATATCCATATCTTGTGTTTAAAGCCCA 888
 QY 1134 CTGAGAAATTTGTCAACACGATTTATTTTGAAGCGCAAAATTTCTTTTGGAGCTGTG 1193
 Db 889 CAGAGAAATTTGGCTGTACAGATCTATAGAGAGCCGAAATTTTCTACCGATCTAGAG 948
 QY 1194 TAAGAGCTGTGTATATATGAGGGAACCCAGCTGGACATTCATTCGCAAAATATAC 1253
 Db 949 TTGTCCTTGTGTATATGATGCTGTATATGATGATGATGATGATGATGATGATGATG 1008
 QY 1254 AAGCTGTATATATATATGATGCTGTATGATGATGATGATGATGATGATGATGATGATG 1313
 Db 1009 GTGATGCCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1068
 QY 1314 AGATTTGCTCAAAACAGATCAAAATCTTATTTTGAAGCTGATGATGATGATGATGATG 1373
 Db 1069 AGATTTGATTAAGATCTTGTGCAAGTCTTATGATGATGATGATGATGATGATGATGATG 1128
 QY 1374 TGGCTTTTGTCCAGAAATGAGAGATTTTCTTCCAGAGATGATGATGATGATGATGATG 1433
 Db 1129 TGGATTTTGAACCTCAATACGCTGTATGATGATGATGATGATGATGATGATGATGATG 1188
 QY 1434 AGGCGCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1493
 Db 1189 TTGCTACACCATATGATTTGTGCTATCTTTCTTAAGGAAATACAGATCTTGTGCTG 1248
 QY 1494 AGTTTAAAGTCAAAATATCTGTTGTGCTGTGGAACAAGTGGTGAAGCATGTAGAG 1553
 Db 1249 ACTTTTGTGATGAA---TATATCTTTTGTGCTGTGGAACAAGTGGTGTATCTCTGAGA 1305
 QY 1554 ATGTTGACGAGACCGTTCTCCAGTGGCCAGTCTCAAAAAGGAAAGCTGTTGAA 1613
 Db 1306 ACATCAACAGAAAGTGTGTTGGTGTGAAGCTTAAACGCTCTTTCTACTGACCA 1365
 QY 1614 TTCTGCGAAACATAGG---GATGAAAGAACTATGCTTTTGTGAAGTAAAGAAAAAG 1670
 Db 1366 TTTTAAAGTGAACAGGAGGATTCATCTTACTTATGATGTTGTGGAACCAAAAAGGAG 1425
 QY 1671 CAGATTTTACTGCACTTTTCTTTGTCAAGAAAAATATCACTACAAAGTATCCATGCTG 1730
 Db 1426 CAGATTTCCCTGAGAGATTTCTTATACATGAAGATATGCTGTGATGATGATGATGAG 1485
 QY 1731 ATTCGGAACAGAGAGAGCGGAGCAAGCTCTTGAAGATTTTCTGTTTGAAGTGGCCAG 1790
 Db 1486 ACCGCTCAGAGAGATGAGAGAGAGCGCTTACACAGTTTCTGCTCAGGAAAAAGCCCA 1545
 QY 1791 TTTTGTGTCTACTTCACTAGCTGCGAGAGGCTGTGATTTGAAGTGAAGTGAAGTGA 1850
 Db 1546 TTTTAAAGTGAACAGGAGTGTGCGAGACAGAGATGATGATGATGATGATGATGATG 1605
 QY 1851 TCAATTTTGAATCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1910
 Db 1606 TCAATTTTGAATTTTGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1665
 QY 1911 GTTGTGGAATATCTGAGAGCAATTCCTTTTGAATCTTGAATCTGGAATCAATTTAG 1970
 Db 1666 GTGTGGAACCTGTGGCTTGTGCACTCATCTTTAA---TGAAAAAATATGAAATATTA 1722
 QY 1971 CACAGCCCTTATTAAGATATGAGATGATGATGATGATGATGATGATGATGATGATGATG 2030
 Db 1723 CAAAGGATTTGTGATCTTCTTGTGAAGCTTAAACAGAGAGTGTCTTCTTGTGGA 1782
 QY 2031 AAATGCTTTATGATACATATCTTGTGCTTCAAGTGTAG 2070

Db 1783 ATATGCTTATGAACACCATCAAGAGGTGCAAGTGTG 1822

RESULT 6
 US-09-764-891-1170
 ; Sequence 1170, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1170
 ; LENGTH: 591
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-891-1170

Query Match 16.0%; Score 356; DB 10; Length 591;
 Best Local Similarity 99.4%; Pred. No. 1.7e-90;
 Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1867 TTCTACATGATGATATATGATCATGATGATGATGATGATGATGATGATGATGATGATG 1926
 Db 1 TTCTACATGATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
 QY 1927 CAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACATTTAGACAGCGCTTATGTA 1986
 Db 61 CAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACATTTAGACAGCGCTTATGTA 120
 QY 1987 AGTATTTGACAGATGCTCAACAGAGATGCTTCCATGATGATGATGATGATGATGATG 2046
 Db 121 AGTATTTGACAGATGCTCAACAGAGATGCTTCCATGATGATGATGATGATGATGATG 180
 QY 2047 ATACATTTCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 2106
 Db 181 ATACATTTCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 240
 QY 2107 AAAGGCAAGACATTTTGAACACAGCTGGGTTTCTTCTTCAAGCTGCCAATCCAGT 2166
 Db 241 AAAGGCAAGACATTTTGAACACAGCTGGGTTTCTTCTTCAAGCTGCCAATCCAGT 300
 QY 2167 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2224
 Db 301 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 358

RESULT 7
 US-10-032-585-6564
 ; Sequence 6564, Application US/10032585
 ; Publication No. US20030180953A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Terry, Roemer D.
 ; APPLICANT: Bo, Jiang
 ; APPLICANT: Charles, Boone
 ; APPLICANT: Howard, Bussey
 ; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 ; FILE REFERENCE: 10182-005-999
 ; CURRENT APPLICATION NUMBER: US/10/032,585
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 8000
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6564
 ; LENGTH: 2019
 ; TYPE: DNA
 ; ORGANISM: Candida albicans
 US-10-032-585-6564

Db	1502	CCGTTGCGCGAGAGGTTTGGATATTCCTAATGTTTCCACGTTATCACTACGATTTAC	1561
Qy	1866	CTTACCATGATGAATGATGTTTCATGCAATTTGGGCGTACTGCTGTGCGAATACAG	1925
Db	1552	CTAGGATATGATGACTACTGTTTCATGTAATTTGTCGTACTGCTGTGCGGTAAGCTTG	1621
Qy	1926	GGAGGCAATTTCTTTTGTGATCTTGAATCGAATACCAATTTAGACAGCCTTAGTAA	1985
Db	1622	GTATTCGCCACTGCTTTTTCGAA--CAGAAACAACAAGATGTTTAAAGATTAATG	1678
Qy	1986	AAGTATGACAGATGCTCAACAGGATGTTCTGTCAGTGTGGAAGAATTTGCC	2038
Db	1679	AATTTATTTCTGAAGCCATCAAGAAAGTCCCAAGCTTTTAACTAATAATTTGCC	1731
RESULT 8			
	US-10-437-963-58448		
	Sequence 58448	Application US/10437963	
	Publication No. US2004012343A1		
	GENERAL INFORMATION:		
	APPLICANT: La Rosa, Thomas J.		
	APPLICANT: Kovacic, David K.		
	APPLICANT: Zhou, Yihua		
	APPLICANT: Cao, Yongwei		
	APPLICANT: Wu, Jie		
	APPLICANT: Boukharov, Andrey A.		
	APPLICANT: Barabak, Brad		
	APPLICANT: Li, Ping		
	TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With		
	TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement		
	FILE REFERENCE: 38-21(53221)B		
	CURRENT APPLICATION NUMBER: US/10/437,963		
	CURRENT FILING DATE: 2003-05-14		
	NUMBER OF SEQ ID NOS: 204966		
	SEQ ID NO 58448		
	LENGTH: 2362		
	TYPE: DNA		
	ORGANISM: Oryza sativa		
	FEATURE:		
	NAME/KEY: unsure		
	LOCATION: (1)..(2362)		
	OTHER INFORMATION: unsure at all n locations		
	FEATURE:		
	OTHER INFORMATION: Clone ID: PAT_MRP4530_60163C.1		
	US-10-437-963-58448		
Query Match	13.6%	Score 301.6	DB 17; Length 2362;
Best Local Similarity	54.6%	Pred. No. 1.5e-74;	
Matches 726;	Conservative 0;	Mismatches 574;	Indels 30; Gaps 5;
Qy	801	AGACAGGCAATTAATCTTCGACAAATACGACACATTCCTGTGGAAATGCTGTGGACATGATG	860
Db	676	ACACTGGCAATCACTTGAATGCTATGAGAGACATTCCTGTGAAACAAGATGGCCATGATG	735
Qy	861	CACCAACGACATTCGACTTTGGAAGAACTAATCTCTGTGAGCACTGAATACAA	920
Db	736	TGGCAGCGCAGCAATACCTTTGCGAGAAATGATTTGGGTGATGCGCTGAATAGAAC	795
Qy	921	TTGCTAAAGCTGGTTTACTTAAGCTTACCTGTGCAAAAATACAGTATTCCTATCATAC	980
Db	796	TACGAGGTGCAATATGTGAAACCTACGCCAGTGCAGCGGTATGCCATTCATCTTCCA	855
Qy	981	TTGACAGACAGATTTGATGGCTTGCTGCAAAACAGGCTTGGGAAGACTGCGGCTTTTC	1040
Db	856	TTGACAGGCGGAGATCTCATGGCTTGCTGACAGAGGCTTGGAAAGACACCGGCTTTT	915
Qy	1041	TCTTACCAATTTTGGCTCATATGATGCAATGAGAAATACTGCCAGTGTGTTTAAAGGT	1100
Db	916	GTTTCCCTATCATTAATGAGATTAAGCTTAGGCGCACACAGAGGCGAAAGGAGATCG	975
Qy	1101	TGCAGG---AACAGAGTGTATTTATTTAGACCACTACGAAATTTGGTCAACCAATTT	1157

Db 976 GGAATGATACCTCTTGATGATTTATCCCCCACTCGGAAATATCAGTCCAAATTC 1035
QY 1158 ATTTGAAAGCCAGAAAATTTCTTTGGAGCTTGTAAGAGCTGTGTATATATGAGG 1217
Db 1036 ATGAAAGAAAGAAATTTGCATACCAAGCTGTGTAGAGGTGTGTGCTATAGTG 1095
QY 1218 GAACCCAGCTGGGACATTCATTCGACAAATAGTACAAGCTGTATATATATGTGCTA 1277
Db 1096 GAGCAACCAATACATCAACAGTTGAGGAATGGAAAGGCGCTGTAATCTTGAGCA 1155
QY 1278 CTCCTGGAAGCTGATGATATCAATAGCAAAAGAAAGATTTGGTCTCAACAGATCAAT 1337
Db 1156 CTCTTGCTGGTGTATGATCTGCTGAGAGGGCTGAGATATCACTGCAATATGTAGT 1215
QY 1338 ACTTATGTTTGGATGAGTGTATGATGATGATGATGATGATGATGATGATGATGAT 1397
Db 1216 ACTTATGCTGTGATGAGTGTATGATGATGATGATGATGATGATGATGATGATGAT 1275
QY 1398 AGTTAATTTCTTGCCCGAGATGCTATCAAAAGAAAGAGCCCAACCTTATGTGATG 1457
Db 1276 AAATTTGTTGAGCAGATGAGATGAGCCCTCGCGGTGAGGCAAGCAATGTTTATG 1335
QY 1458 CAATTTTCCAGAGAAATTCAAAGTTGGCTGAGAGTTTAAAGTCAATATATCTG 1517
Db 1336 CTACTTTTCCAAAGAAATACAGGCGATGCTTCAG---TTTCTTGCGGACTACATCT 1392
QY 1518 TTGTTGCTGTGACAAAGTGGGTGAGCAGATGATGATGATGATGATGATGATGATG 1577
Db 1393 TCCTTGTGTAGAGAGTGTGTGATGATGATGATGATGATGATGATGATGATGATG 1452
QY 1578 TTGGCCAGTCTCAAAAGAAAGAAAGCTCGTGAATTTGCG-----G 1620
Db 1453 TCCTTGAAGCGGATTAAGAAAGTAACTGATGATGATGATGATGATGATGATGATG 1512
QY 1621 AAATGAGGAGATGAAAG---ACTATGCTTTTGTGAACTAAGAAAGAAAGCAAT 1676
Db 1513 GTACTTCAATGGAACCAAGCGCTTACTCTGCTTTGTGAGACAAAGAGGAGATGATG 1572
QY 1677 TTAATGCAATTTCTTTGTGACAAAGAAATATCAATCAATGATGATGATGATGATG 1736
Db 1573 CTTTGGAACTGCTTTTACCAAAAGCGGTTCCCTGACAAAGATTCATGAGGACCG 1632
QY 1737 AACAGAGAGCGGAGCAAGCTCTTGAAGTTTCCTTTGGAAGTGGCCAGTTCTTG 1796
Db 1633 CACAGAGAGAAAGAGATGAGCGCTTATGATCATTCAGAGTGGAGCAACCCATCTTG 1692
QY 1797 TTGCTTACTTCAATGCTGCGAGAGGCTGATATGAAATGTCGAACATGTTATCAAT 1856
Db 1693 TTGGAGCTGATGATGCTGCGCGGAGCTTGAATGATGATGATGATGATGATGATG 1752
QY 1857 TTGATCTTCTTACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1916
Db 1753 TTGATCTTCTTACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1812
QY 1917 GGAATGCTGAGAGCAATTTCTTTTGAATCTTGAATGATGATGATGATGATGATG 1976
Db 1813 GGAATGCTGAGAGCAATTTCTTTTGAATCTTGAATGATGATGATGATGATGATG 1869
QY 1977 CTCTAGTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2036
Db 1870 CTCTAGTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1929
QY 2037 CCTTATGATCAATCTTCTGCTTCACTGATGATGATGATGATGATGATGATGATG 2096
Db 1930 CCGGCGGCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1989
QY 2097 TTGATACAG 2106
Db 1990 TTGGTGGCG 1999

Sequence 46077, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xiandeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 46077
LENGTH: 1815
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-46077
Query Match 13.4%; Score 298.2; DB 16; Length 1815;
Best Local Similarity 54.2%; Pred. No. 1.2e-73;
Matches 684; Conservative 0; Mismatches 553; Indels 24; Gaps 3;
QY 804 CAGGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 863
Db 353 CTGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 412
QY 864 CAGGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 923
Db 413 CTGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 472
QY 924 CTAAGCTGTTAATCTTAAGCTTACTCTGTGCAAAATATAGATTTCCATATCTTG 983
Db 473 AATGAGCGCTTCAACCAAGCAAGCAAGCTGCAAAATATAGCTTCCATATCTTG 532
QY 984 CAGGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 1043
Db 533 CAGGATTAATCTTGACAAATAGCACTATTTCTGTGAAAGTGTGCAATGATGAC 592
QY 1044 TACCAATTTGCTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1085
Db 593 TCCAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 652
QY 1086 GTGCTTTTAAAGTGTGACAGAAACCAAGTATTTTGTAGACCACTGAGAAATGG 1145
Db 653 CTTTATCCAAAGAAAGGCTTACCAAGCTGCTGATTAATGCTCACTAGAGGTTGG 712
QY 1146 TCAACAGATTAATTTGAGAGGCAAGAAATTTCTTTTGGAGCTGTGATGAGCTG 1205
Db 713 CCAACCAATTTTGCATGAGAGCAAGAAATTTTACTTATAGATCCGAGGCTGAG 772
QY 1206 TTAATATGAGGAAACCAAGCTGAGCAATTCATTCAGAAATATGACAGGCTGAT 1265
Db 773 TCGTACAGGAGTGTCTCCAAATGATGATGATGATGATGATGATGATGATGATG 832
QY 1266 TATTAATGCTCTCTGGAAGAGTATGATGATGATGATGATGATGATGATGATGAT 1325
Db 833 TTTTATGCTCTCTGGAAGAGTATGATGATGATGATGATGATGATGATGATGATG 892
QY 1326 AACAGATCAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1385
Db 893 CCAAGCTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 952
QY 1386 CAGAAATGAAGATTAATTTTTCGCCAGAAATCCATCAAGAGAGAGCGCCAAACCC 1445
Db 953 CTCAAATTAAGATTAATTTTTCGCCAGAAATCCATCAAGAGAGAGCGCCAAACCC 1012
QY 1446 TTAATGCTCAAGTATTTTCCAGAGAAATCAAGGTTGCTGAGAGTATTAAGT 1505
Db 1013 TGAATGCTCAAGTATTTTCCAGAGAAATCAAGGTTGCTGAGAGTATTAAGT 1069

QY 1506 CAATATTACTGTTTGTGCTGTGTGGACAAGTGGGTGGAGACATGTAGACATGTTCAGACA 1566

Db 1070 GTGACTCAATCTTTTGTGTGTGTGGAGTGGTCTAATTCAGAAAAACAATTACTCAA 1123

OY 1566 CGGTCTCCAAAGTTGGCCAGTCTCTCAAAAAGAAACAGCTCGTTGAATTCGGCAACA 1626

Db 1130 AAGTCTTACGTTTAAATAATCAAGATTAGAGATCAGCCTTATTSATCTATTGTGCAT 1183

OY 1636 TAGGGGATGMAAGAACTATGCTCTTTGTTGAACCTAAGAAAAAAGCAGATTTTACTGCA 1686

Db 1190 CCACTGACGGTTTGACTTTGATCTTTGTGAAAACTAAGAAATGCGACATCAATTGACCG 1243

OY 1686 CTTTCTCTTTCACAAAAAATAATATCAATCAAGATTCATAGTGTATGGGAAACGAGAG 1745

Db 1250 ATTCTCTATCATGCAAAACTTTAGAGTACCGCCATTCATGAGACCGTATCCCAATCTG 1309

OY 1746 AGCGGGACCAAGCTCTTGAGATTTTCCTTTGGAAAGTGCCCACTTCTGTGCTACTT 1805

Db 1310 AGAGAGAACGTGCTTGGCCGCTTCAATCTGGTCCGCTACTTATTGGTTGGACAG 1363

OY 1806 CAGTAGCTGCAGAGGCGCTGGAATTTGAAATGTGCAACATGTTATCAATTTGATCTTC 1865

Db 1370 CTGTGCGACGTAGAGGCTAGATTAATTCAAAACGTCAACCAGGTTATCAACTACGATTTAC 1429

OY 1866 CTCTACCAATTGATGAATAATGTTCATGCAATTTGGCGGATACGTGCTGTGGAAATATGT 1925

Db 1430 CAAGTATGTGATGATTAATCGTCCATTAATAATTTGGTAGACGTGTGTCGGGTAAACACCG 1489

OY 1936 GCAGAGCAATTTCCTTTGTGATCTTGAAATCGGATTAACATTATAGCAAGCCTCACTAA 1985

Db 1430 GTCTTGCACTGCTTTTTCAAAGATGAAAAACAGTAAC--ATTGTAAAGGTTTGCTATG 1546

OY 1986 AAGTATGACAGATGCTCAACAGATGTTCTGCAATGTTGGAAAGAAATTCCTTTAGTA 2045

Db 1547 AAATTTTACAGAGCTAACCAAGAGATCCCATCATCTTTGAAGGACCTATGATGAGTG 1606

OY 2046 C 2046

Db 1607 C 1607

RESULT 10

US-10-369-493-46167
 ? Sequence 46167, Application US/10369493
 ? Publication No. US20030233675A1
 ? GENERAL INFORMATION:
 ? APPLICANT: Cao, Yongwei
 ? APPLICANT: Hinkle, Gregory J.
 ? APPLICANT: Slater, Steven C.
 ? APPLICANT: Goldman, Barry S.
 ? APPLICANT: Chen, Xianning
 ? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ? FILE REFERENCE: 38-10(52052)B
 ? CURRENT APPLICATION NUMBER: US/10/369,493
 ? PRIOR FILING DATE: 2003-02-28
 ? PRIOR FILING DATE: 2002-02-21
 ? NUMBER OF SEQ. ID NOS: 47374
 ? SEQ ID NO 46167
 ? LENGTH: 1854
 ? TYPE: DNA
 ? ORGANISM: Saccharomyces cerevisiae
 ? US-10-369-493-46167

Query Match	13.4%;	Score 298;	DB 16;	Length 1854;
Best Local Similarity	54.5%;	Pred. No. 1.4e-73;		
Matches 676;	Conservative	0;	Mismatches 540;	Indels 24;
				Gaps 3

Qy 804 CAGGATAACTTCGCAAAATCGACCTATTTCTGGAGGTGTGAGCATGATGCAC 863

Db 389 CTGGATAAAGTTTGATTAATTACGATAATATTTCCCGAGATGCATCTGGAAAAGATGTC 448

QY	864	CACCGCAAACTCTGACTTTTGAAAGAGCTAATCTCTGACACACACTGATATTAACAACATCTG	923
Db	449	CTGAACCTATCTAGATTTTAGTTCTCCCTCTCGAGTACCTGTATATGGAAAAATATCA	508
QY	924	CTAAAGCTGGTATATACAGCTTACTCCTGTGCAAAAAATACAGATATTCATATCATCTTG	983
Db	509	AACCTGGTAGTTCCACAAACCTACTCCAGTACAAAAATATTCATATCCAAATAGTTATCCA	568
QY	984	CAGACACGATTTTATATGGCTGTGCTGCTCAAAAGGGCTCGGGAAGACTGCGGCTTTCTCC	1043
Db	569	AAGTAGGGAATTTATGTCATGCGCACAAAGGGCTCAGGTAGACAGGTGGTTTTGT	628
QY	1044	TACCAATTTTGCTCATATGATGATGATG-----AATTAATGCCA	1085
Db	629	TTCCGCTCTACGAGAGCTATATTCAGAGTGTGGGCCCCCTCCAGTACCCGAGAAAGACAGA	688
QY	1086	GTCGTTTTAAAGATGTGACGAGACACAGATGTATATTTGTAGCAACCACTGAGAAATTTG	1143
Db	669	GCTTTATCTTGAAGAAAGGATATTCATGCTGTTAGTACTTGACCAACTAGAAATTTG	748
QY	1146	TCAACAGATTTATTTGGAAAGCCAAAAATTTCTTTGGGACTTGTTGTAAGCTGTTG	1205
Db	749	CTACCAAAATTTTGAAGAAACAAAGAAATTTAGGTATAGTCTTGGGTTTGGCCATGCG	808
QY	1206	TTATATATGGGGGAAACCCAGCTGGGACATTCATTTCCAAATATGTATCAAGAGCTGTATA	1265
Db	809	TTGTTTATGTGCGCGCCCAATCGGTACCGATGAGAGAGGTTGACCGCGCTGTACT	868
QY	1266	TATTATGTCTACTCTCGGAAAGCTGATGATGATCATATGSCAAAGAAATGCTCTCA	1325
Db	869	TATTGTGCTACACAGGGGCGGCTGATATGACTTATTAAGACGGGAAAGATTTCTTTGG	928
QY	1326	AACGATCAAAATCTTAGTTTGTGATGAAGCTGATCGATTTTGAACTATGCGTTTGGTC	1385
Db	929	CTATATCAATTAATCCTGTTATGATGAGCTGTATAGATCTGATATATGGGATTTCCAGC	988
QY	1386	CAGAAATGAGAGTTAATTTCTTGCCAGAAATGCCATCAAGGAAACGGCCAAATCC	1445
Db	989	CCCAAAATCGCCATATATGTTGAAAGATGTGATAGCCCTCTGTGAAATATGAACAGACT	1048
QY	1446	TTATGTCAGTGAACATTTTCCAGAGAAATTCAAAGGTGTGGCTGCAGAGTTTAAAGT	1505
Db	1049	TGATGTTTCCGCCAATTCCTCATATGATATCCAGCATTTGGCAGCGGATTTCTTAG--	1106
QY	1506	CAATATATCTGTTGTTGCTGTGGAACAATGGGTGAGCATATGAGATGTTACAGAGA	1565
Db	1107	-TAATATATTTTATTTTTCAGATAGAAAGAGTCGGTTCCACTCGAAAAATATCAAGCAA	1165
QY	1566	CCGTTCTTCCAGTTGGCCAGTGTCTCAAAAAGAGAAAAAGCTGTGAAATTTTCGAAACA	1625
Db	1166	GAATATATATGTCGATGACATGACAAAAAGTGGGTTGTATGATTTACTGCTGAG	1225
QY	1626	TAGGGGATGAAGAAGCTATGCTCTTTGTTGAAATTAAGAAAAAAGCAAGTTTACTGCA	1685
Db	1226	AGCATATAGGGATGACATGATGATTTTGTGAAAGAAAAAGATAGCGGATCTAACCTCAG	1285
QY	1686	CTTTTCTTTGCAAGAAAAATATCAACTACCAATATCCATGCTGATCGGAAACAGAG	1745
Db	1286	ATTTTGTGATCATCAAAATTTTCAAAGCTACAGCATATATGAGACCGCACAGGCTG	1345
QY	1746	AGCGGAGCAAGCTCTTGAGATTTTGGCTTTGSAAGTCCCAAGTCTTGTTGCTACTT	1805
Db	1346	AACGTGAACGGCTTATCTGCTTTCAAAGCTAACGTAGCTGATATCTGCTGCCAACAG	1405
QY	1806	CAGTAGCTGCAGAGGCTGGATTTTGAATAATGTGCAACATGTTATCAATTTTGATCTTC	1865
Db	1406	CTGTAGACGAGAGGTTTGGATATACGAGATGTACACATATGTCATTATATGATTTGC	1465
QY	1866	CTTCTACATTTGATGATATGTTATGATGAATTTGGCGTACTGCTGTGTTGREGAATACTG	1925
Db	1466	CTCTCGACATGAGACTTACGTTTCCAGAAATGTTGTAACACAGGCTGCTGTGTATCCAG	1525

Db 1466 C C T C C G A C A T C G A T G A C T A C G T T C A C A G A A T T G T A G A A C A G C C C G T G C T G G T A A C A C A C A G 1525

QY 1926 GCAGACAAATTCCTTTTGAATCTTGAATCGGATACCATTTAGACACCCCTTAGTAA 1985
DB 1526 GCCTGCTACTTATCTTCAATAGTAA---TATCAGAAATATGGAAGGTTGATGG 1582
QY 1986 AAGTATTTGACAGATGCTCAACAGAGATGTTCTGCATGTT 2025
DB 1583 AATTTTAAATGAGGAAATCAAGAGTTCCCAAGTTT 1622

RESULT 11
US-10-369-46378
; Sequence 46378, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46378
; LENGTH: 1908
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46378

Query Match 13.2%; Score 294.6; DB 16; Length 1908;
Best Local Similarity 56.0%; Pred. No. 1.3e-72;
Matches 661; Conservative 0; Mismatches 484; Indels 36; Gaps 4;

QY 803 ACAGGCAATTAACCTTCGCAAAATAGACATATTTCTTGGAAGTGTGACATGATGCA 862
DB 436 ACTGGATTTAATCTCGAAAGATGATGACATCCCTGTATAAGTTCCGGTGAATATC 495
QY 863 CCACGACAAATTCGACTTTGGAAGACTAATCTGTGACACATGAAATTAACAAT 922
DB 496 GAGCTGTATAGTTCACCTTCCCTCTTGAACCTCA---TCTTCTCAAAACAT 552

QY 923 GCTAAGCTGTATTAATTAAGCTTAATCTGTGCAAAATACATATTCCTATCAACT 982
DB 553 AACTCTCTGCTACACTCAGCCCAACCCAGTTCAAAAGAACTCATTCCTATTTGACA 612
QY 983 GCAGAGAGATTTGATGAGGCTTGCTCAACAGAGTCTGGGAAGAGTGGGCTTTCTC 1042
DB 613 TCCGACCTGATTTGATGAGCTGTGCCCCAACCAGTTCCGTTAAGATCTGCTTTCTTA 672

QY 1043 CTACCAATTTGGCTCATATGATGATGATGAAATTAATGCAAGTGTGTTTAAAGTTG 1102
DB 673 TTCCCTATTTCTTCTGCTTTGATTAAGACCTCCCGCTGTCTCGTTGATCAGGAT 732
QY 1103 CAGG-----AACCAAGTATATTATTGAGCAGCACT 1135
DB 733 GCCGGTATGAGTTATGCTCTCGCAAGACCTATCTTACCACTTATTTGGCTCCACT 792

QY 1136 CGAGATTTGCTCAACAGATTTATTTGAGCCAGAAATTTCTTTTGGACTTGTGA 1195
DB 793 CGTGAATAGTTTGTCAATTCATGAGGAATCTGTAATCTGTACCGTTCTTGGGTG 852

QY 1196 AGAGCTGTTGTTATATATGAGGGAACCCAGCTGGACATTTAATTCGCAAAATGTAACA 1255
DB 853 CGTCCCTGTGCTGTTTAAAGGAGTCCGATATGCTGTCAAAATCCGCAATTTGACCAA 912

QY 1256 GGCTGTAATATATATATGCTACTCTGGAAGACTGATGATATATGCAAAAGAAAG 1315
DB 913 GGTTCGACTTCTTCTGCTACTCTGCTGTGTTGTTGTTATTTAATGATCGTGTGT 972

QY 1316 ATGCTCTCAAAACATCAATTAATCTTGTGATGATGATGATGATGATGATGATGATG 1375
DB 973 ATTCTTTGGCAACATTAATTTTGTGTGTGATGATGATGATGATGATGATGATGATG 1032

QY 1376 GGTTTTGGTCCAGAAATGAGAAATTAATTTCTTGCCAGAGATGATGATGATGATGATG 1435
DB 1033 GGTTCAGCTCTCAATTTGTCATATGTCATATGTCATATGTCATATGTCATATGTCAT 1092

QY 1436 CGCAAACTTATGTTCAATGTCATATGTCATATGTCATATGTCATATGTCATATGTCAT 1495
DB 1093 CGTCAAACTCTCAATGTTTTCGACATTTCCCGCTGATATGTCATATGTCATATGTCAT 1152

QY 1496 TTTTAAAGTCAATTTATGCTGTTGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1555
DB 1153 TTTTAAAG---GATATGCTTTCTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1209

QY 1556 GTTCAGACACCGTTCCTCAAGTTGCGCAGTTCTCAAAAGAAAGAAAGCTGTTGAAT 1615
DB 1210 ATTACTCAAAAGTGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1269

QY 1616 CTGCAAACTATGAGGATGAAAG---AATATGCTTTGTCATGTCATGTCATGTCATGTCAT 1672
DB 1270 CTTCACACCTTACCTCTGAGGCTTACCTGATCTTTGATGATGTCATGTCATGTCAT 1329

QY 1673 GATTTTACTGCAACTTTCTTTGTCAGAAATTAATCAATGTCATGTCATGTCATGTCAT 1732
DB 1330 GACACACTTACCGACTATTTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1389

QY 1733 CGGGAACAGAGAGCGGAGGAGGCTTTGAGATTTTGGCTTTGCAAAAGTCCCACT 1792
DB 1390 CGTACCAACGTCAGGCTGAGGCGCTTTGATGTCATGTCATGTCATGTCATGTCAT 1449

QY 1793 CTGTTGCTACTCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1852
DB 1450 ATGTCGCTACTGCGTGCAGTGCAGTGTGTCATGTCATGTCATGTCATGTCATGTCAT 1509

QY 1853 AATTTGATCTTCTTCAACATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1912
DB 1510 AACTAGATTTTACCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1569

QY 1913 TGTGGATATCTGTCAGGCAATTTCTTTTGTGATCTGTA 1953
DB 1570 GGTGTAACTGTCAGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1610

RESULT 12
US-10-424-599-14656
; Sequence 14656, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 14656
; LENGTH: 3012
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113240C.1
US-10-424-599-14656

Query Match 13.1%; Score 291.4; DB 13; Length 3012;
Best Local Similarity 54.3%; Pred. No. 1.5e-71;
Matches 719; Conservative 0; Mismatches 566; Indels 38; Gaps 5;

QY 769 TGAGATGAGACTCATCTTTGACATATATGACAGGCAATTAATCTGACAAATATGCA 828

Db 749 TAAAGCTGAAGAGGCACTTCACTGAGCAGAGAAATACAGAAATTAACCTTTAATGCAATGTA 808
 Qy 829 CACTATTTCTTGTGAAGTGTCTGACATGATGACCAACCAATTTGTAATTTGTAAGA 888
 Db 809 AGGCACTTGAAGTGAACCAAGTGTGACAAATGTCCCACTGTGAAATACGTTTGAGA 868
 Qy 889 AGCTAATCTGTGAGACACTGATAACACATTTGCTAAAGCTGTGTAATCTAAGCTTAC 948
 Db 869 GATTGATTTGGGTGAAGCACTTAATCAGAAATTAAGAAAGTGCAGAAATGTGAAGCAAC 928
 Qy 949 TCCGTGCAAAAATACAGTATTTCTATCATCTTGACAGACAGATTTGATGGCTTGTG 1008
 Db 929 ACCGTGTACGGGCGATCCATACCGATATCTGTGAGACGGGATTTGATGGCTTGTG 988
 Qy 1009 TCAACAGAGGTCTGGGAAAGATGGGGCTTTCTCCATCAATTTGGCTCATATGATGCA 1068
 Db 989 GCACATCTGTGTGGAAGAGACGTGCATTTCTGCTCCGATTAATAGTGAATCATATGAG 1048
 Qy 1069 TGATGAATTAATGCGCATGCTGTTTAAAGATTGCG-----AGGAACAGAGGTATAT 1122
 Db 1049 GGGCCAAACCCCTGACAGAGGCACTCGGGGGTGGCGTACAGTGTATCCACTTGGCGTTGT 1108
 Qy 1123 TGTAGCAACCACTGAGAAATTTGCTCAACGATTTATTTGGAAGCCAGAAATTTTCTTT 1182
 Db 1109 TCTTTCACCAACGAGGAGGCTATCCATGCAAAATACATGAAGAGCTTGAAGATTTTCATA 1168
 Qy 1183 TGGGACTGTGTGAAGAGCTGTTGTTATATATGGGGGAACCCAGCTGGGCACTTCAATTTG 1242
 Db 1169 CCAACCTGGGGTTAAGGTGTGTGTTGCTATATGCTG-----ACCAACAAATTAACCAACGAG 1223
 Qy 1243 ACAAAATGTACAAAGCTGTAAATATATATGCTGTACTCTGTGAAGACTGATGATATCAT 1302
 Db 1224 GGAATCTTGAAGAGGGGTGAGCACTTCTGTGCACTCTGGAAGACTGTGATTTGCT 1283
 Qy 1303 AGGCAAGAAAGAAAGTGTGCTCAACAGATCAAACTCTAGTTTGGAGTGAAGCTGTATG 1362
 Db 1284 GGAAGAGCTAGAGTTTCACTGCAATGATTCGATCTGCGCTTATGATGAGCGAGATG 1343
 Qy 1363 CATGTGTGATATGGGTTTGTGTCAGAAAATGAAGATTAATTTCTTCCAGAGATGCG 1422
 Db 1344 GATGCTGATATGGGTTTGTGAGCAAAATGAAGAAATTTGTAAGCAATGACATGCG 1403
 Qy 1423 ATGAAGGAAGAGCGCCAAACCTTATGTTCACTGCACTTTTCCAGAGAAATTCGAAAG 1482
 Db 1404 TCCACAGAGTGCAGACAGACTATGTGTTGCTAGTGAACATTTCCAAAAGAAATTCAGAG 1463
 Qy 1483 GTTGGCTGAGAGTTTTAAAGTCAATATATCTGTTGTGCTGTGGAAGAATGGAGTGG 1542
 Db 1464 ATGGCTTCTGA-----TTTCTTTCAAAATTAATTTTCTGCTGTGGAAGTGGAGTTC 1520
 Qy 1543 AGCATGTAGAGATGTTCAAGACCGTTCCTCAAGTTGGCCAGTTCTCAAAAAGAAAAA 1602
 Db 1521 AAGTACTGATTAATATGTCAAAAGAGTGTAGTATGTTCAAGAGTCTGACAAAGAAATGCA 1580
 Qy 1603 GCTCGTTGAAATTTCTGCG-----AAACATTAAGGGGATGAAGAAAGAC 1641
 Db 1581 CCTAATGACCTTTCTTCATGACAGAGGGCAATGTGTGACAGAAAGCAAGCTTTAAC 1640
 Qy 1642 TATGCTTTGTTGAACTTAAGAAAAGACAGATTTTACTGCAACTTTTCTTTGTCAAGA 1701
 Db 1641 TTTAGTTTTTGTGAGACTAATAGAGAGTGTATGCTGGAACACTGTTGTGTAA 1700
 Qy 1702 AAAAATATCACTACAGATCCATGTGATGAGGGAACAGAGAGCGGGAGCAAGCTCT 1761
 Db 1701 TGGTTTTCCGGCAACTACTATGATGAGAGACAGGTCACAGAGGAAAGAAATTAAGCTTT 1760
 Qy 1762 TGAAGATTTTCCCTTTGGAAGTGCAGATTTCTTTGTGCTAATTTAGTAGTGCAGAG 1821
 Db 1761 GAGTCATTTAAAGTGCACACCCCAATATGTGTCACACGATGTGTGACAGCTGG 1820
 Qy 1822 GCTGATATTTGAAGATGCAACATGTATCAATTTGATCTTCTTTACCATTTGATGA 1881

Db 1821 TCTTAATATCCCATGTTGCTCATGTGTCACATTTGACCTTGCCAAATGACATTTGATGA 1880
 Qy 1882 ATATGTTCACTGCAATTTGGGCGTACTGTGTGTTGGGAATATCTGCAAGCAATTTCTT 1941
 Db 1881 TTATGTACACCGAATTTGAAGAACAGGCGAGCTGGAAGAAAGAGTCTTGAATCTGCAATT 1940
 Qy 1942 TTTTATCTTGAATGTGATTAACCATTTTGCACAGCTCTAGTAAAGATTTGACAGATGC 2001
 Db 1941 CTTTAA--TGACAACAATTCATGCTGCTAGCTAGAGCTTTATCAGAACTGATGCAAGAGC 1997
 Qy 2002 TCAACAGATGTTCCGTGATGATTTGGAAGAAATTTGCTTTAGTACATATCTTCCGCTT 2061
 Db 1998 AAATCAAAAGATCTGCTGTGCTCTGCTCAGCGTTTGCTGCTGATCTTTTGTGTGAGG 2057
 Qy 2062 CAG 2064
 Db 2058 CAG 2060
 RESULT 13
 US-10-424-599-14660
 ; Sequence 14660, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 14660
 ; LENGTH: 2672
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_113244C.1
 ; US-10-424-599-14660
 Query Match 12.7%; Score 282.2; DB 13; Length 2672;
 Best Local Similarity 54.1%; Pred. No. 5,7e-69;
 Matches 737; Conservative 0; Mismatches 578; Indels 48; Gaps 6;
 Qy 769 TGAGATGAGAGCTCCATCTTTGCACTTATTCAGACAGGATTAACCTTGCAAAATACGA 828
 Db 537 TAAAGCTGAAGAGGCACTTCACTGAGAGGAAATTAACGAAATTAATTTGATGCAATGTA 596
 Qy 829 CACTATTTCTTGTGAAGTGTCTGACATGATGACCAACAGCAATTTCTGATTTGGAAGA 888
 Db 597 AGACATTCAGTGAAGACAGTGTGACAAATGTGCCCACTGTGAATACGTTTGCAAG 656
 Qy 889 AGCTAATCTGTGACAGCTCTGAATTAACAACATTTGCTAAAGCTGTGTAATCACTTAC 948
 Db 657 GATTGATTTGGGTGAAGCACTTAATCAAGATTAAGAAAGTGCAGAAATATGTAAGCAAC 716
 Qy 949 TCCGTGCAAAAATATACAGTATTTCTATCAATTTGAGAGACAGATTTGATGGCTTGTG 1008
 Db 717 ACCGTGTACGGGCAATGCCATACCATATCTGAGTGAACGGGATTTATGCTTGTG 776
 Qy 1009 TCAACAGAGGTCTGGGAAGACTGGGCTTTTCTCTCAACAATTTTGGCTCATATGATGCA 1068
 Db 777 GCAGACTGTGTTGGAAGACAGCTGCATTTCTGCTCCGATTAATCAGTGAATCATGAG 836
 Qy 1069 TGATGAATTAATGCGCATGCTGTTTAAAGATTGCG-----AGGAACAGAGGTATAT 1122
 Db 837 GGGCCAAACCTGTGACAGAGGCACTGTGGGGTGTGTAAGTGTATCCCTTTCGCTTGT 896
 Qy 1123 TGTAGACCAACTGCGAGAAATGGTCAACAGATTTATTTTGAAGCCAGAAAATTTTCTTT 1182
 Db 897 TCTTCTCCAAAGAGGAGTATCAATGCAATATCAATGAAGAGGCTTGAAGATTTTCATA 956

OY	1183	TGGGACTTGTGTAGAGCTGTGTGTATATATGSGGGAAC-----CCAGCTGGGAC	1233
Db	957	CCAACTGGGGTTTGGGTAGTTGTTGTGATATGTGTGGAGCACCAATTAACGACGAGGAGA	1016
OY	1234	TTCA-----ATTGACAAATAGTACAAAGCTGTATATATATATGTCTACTCTCTGAAG	1287
Db	1017	TGCAGGGACACTTACACAGTGGCTTGGGGGTGGACATTTCTGTGCAACTCTCTGGAG	1076
OY	1288	ACTGATGGATATCTATAGGCAAAAGAAAGATTGCTCTCAACAGATCAAACTACTAGTTT	1347
Db	1077	ACTGTGATTTTGTCTGGAGAGCTTGAAGTTTACTGTCAATGATTTGGAATPCTGGATT	1136
OY	1348	GGATGAAGCTGATGCAATGTTGATATGGTTTTTGTCCAGAAATGAAGAAATTTC	1407
Db	1137	AGATGAGGCAATGCGATGCTGATATGGGTTTTGAGCCCAATTAAGAAAGATTGAGA	1196
OY	1408	TTGCCCCAGATTTGCCATTCAAAGGAACAGCCCAACCCCTATGTTCAAGTCACTTTCC	1467
Db	1197	GCAATATGACATGCTCCACCAAGTCCACAGACACTATGTTGTTCACTGCACATTTCC	1256
OY	1468	AGAGAAATTCAAAGTTGGCTGCGAGATTTTAAAGTCAATATCTGTTTGTCTGT	1527
Db	1257	AAAGAGATACAGAGATTGCTTGGA-----TTTCTTCAATATATATTTCTTGTCTGT	1313
OY	1528	TGGACAAATGGGTGGAGCATGTAGAGATTTTACAGACCGTTCTCCAGTTGGCCAGTT	1587
Db	1314	TGGAGAGATGGGTTCAAGTACTGATTTATTTATGTCCAAAGATTGATGATTTCAAGATC	1373
OY	1588	CTCAAAAAGAAAGACTCGTTGGAATTTCTGCG-----AAACAT	1626
Db	1374	TGACAGAGAGATGACCTATAGACCTTCTCATGACACAGGGCAATGGTGTACAGG	1433
OY	1627	AGGGGATCAAAAGAACTATGTTCTTTGTTGAACATTAAGAAAAAGCAATTTTACTGCAAC	1686
Db	1434	AAAGCAAGCTTTAATTATGTTTTTGTGGACACTTAAGAGGAACTGATTCGCTGGAA	1493
OY	1687	TTTTCTTTGTCAAGAAAAAATATCAACTACAAGATTCATGATCGGGAACAGAGAGA	1746
Db	1494	TTGGTTGTCTTAATGTTTTCCTCCAGCACTACATTCATGATGATGACAGGTACACGACGA	1553
OY	1747	GCGGAGCAAGCTCTTGAGATTTTGTGTTGGAAAGTCCCAAGTTCTTGTGCTACTTC	1806
Db	1554	AAAGAAATTAAGCTTGAGATTATTTAAAGTGGCAACCCCCAATTTGGTGGCACAGA	1613
OY	1807	AGTAGCTGCCAGAGGGCTGGAATATTTGAAATATGTGCAACATGTTATCAATTTTGAATCTCC	1866
Db	1614	TGTTGTCTCACTGTGCTTTGATATTTCCCAATGTTCTCATGTGTGCACTTTGACCTGCC	1673
OY	1867	TTCTACCAATTGATGATATGTTTCATCGAATTTGGGGGTACTGTTGTTGTGGGAATACTGG	1926
Db	1674	AAATGACATTGATGATTTATGTACACCGAATTTGAGAGAACAGGGCAGCTGGAAGAAAGG	1733
OY	1927	CAGAGCAATTTCTTTTGTGATCTTGAAATCGGATACCAATTAGACAGCCCTCTAGTAA	1986
Db	1734	TCTTCAACTGCTGTTCTTTAA---TGCACAATATCATCTCTGGCTTAAGCCTTATACGA	1799
OY	1987	AGTATTTGACAGATGCTCAACAGAGATTTCTGCATGTTGTGAGAAATTTGCTTTAGTAC	2046
Db	1791	ACTGATGCAAGAAAGCAAAATCAAGAAATGCTCTGCTGCTTCACAGGTTTGTGCTGCTGC	1850
OY	2047	ATACATCTCTGGCTTCACTGTGATAGTACAAAGAGAAACGTGTTT	2089
Db	1851	TTCTCTTTGTGGGGAAGAACCGTGCATCAGGCGGACCGTTT	1893

	APPLICANT:	Slater, Steven C.
	APPLICANT:	Goldman, Barry S.
	APPLICANT:	Chen, Xianfeng
	TITLE OF INVENTION:	EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
	TITLE OF INVENTION:	PLANTS WITH IMPROVED PROPERTIES
	FILE REFERENCE:	38-10(5205)B
	CURRENT APPLICATION NUMBER:	US/10/369,493
	PRIOR FILING DATE:	2003-02-28
	PRIOR APPLICATION NUMBER:	US 60/360,039
	PRIOR FILING DATE:	2002-02-21
	NUMBER OF SEQ ID NOS:	47374
	SEQ ID NO	26954
	LENGTH:	1681
	TYPE:	DNA
	ORGANISM:	Neurospora crassa
	US-10-369-493-	26954
Query Match	12.4%;	Score 276.2; DB 16; Length 1681;
Best Local Similarity	53.2%;	Pred. No. 2,1e-67;
Matches	673; Conservative	0; Mismatches 563; Indels 29; Gaps 3;
Qy	798	ATCAGACAGGCATPAACTTTCGACCAATACAGACTATTCTTGGAAGTGTGACATG 857
Db	239	AGCACACTGGTATCAACTTCGAGAAATAAGATGACATCTCCCGTTGAGCGTTCCGGTAGCA 298
Qy	858	ATGCACCACCAACAATTCTGACTTTTGAAGAAGCTAATCTGTGACAGACTGAATPACA 917
Db	299	ACGGGCCGAGCCCGTTTCTTAACCTTCAGAACACCTCTCTTGACACACCACTCATTTCCA 358
Qy	918	ACATTTGCTAAGCTGTATTACTTAAGCTTACTCTCTGTCGAAAAATACAGATTTCTATCA 977
Db	359	ACATTTAGCTTCGCCCGCTACAATGTCCCACTCCGGTTCAGAAATGACTCCATCCCACATG 418
Qy	978	TACTTGCAGACAGAGATTGATGGCTTTGGCTTCAAACAGAGTCTGSGAAGAATGGGGCTT 1037
Db	419	TCATGGGTGGCCGGATCTGATGGCTTGGCCGCCAGACTGGTTTCGGCAAGACTGTGGTIT 478
Qy	1038	TTTCTACCAACAATTTTGGCTCATATGATGATGATGGAATPAC----- 1080
Db	479	TCCCTTTCCCCATCTTCTTCAGTCTTTTCCACACTGGTCGATCTCCGATTCGCCGACGCG 538
Qy	1081	-----TGCAGTTCGTTTTAAAGAGTTGCAGAGAACCAAGTATATTATTGACACCAA 1133
Db	539	CTGCTGGTGGCTATAGTTCGTCAAGCGCAAGGCTTACCCCAACCGCTTGATCTTGTCCCA 598
Qy	1134	CTCAGAAATTGCTCAACACCAATTATTTTGAAGCCAGAAATTTTCTTTGGAGCTGTG 1193
Db	599	CCCGTGTGTTGTCTTCAATCTACAGACAGAGCTCGAAGTTCCGCTACCGTTCGGGG 658
Qy	1194	TAAAGCTGTGTTTATATATATGGGGGAAACCAAGCTGGGACATTCAATTTCACAAATATGATAC 1253
Db	659	TCCGTCTTGGTGTGTCTACCGGTGGTGGCCATATTCGGCTTCCAGCTTCGCGACATCGAGC 718
Qy	1254	AAGGCTGAATATATTATNGTCTACTCTCTGSAAGACTGATGATTCATATGAGCAAGAAA 1313
Db	719	GCGGTGGCATCTTCTTGTGTCTACTCTCTGTGTCTGTTCATCTCATGAGACGGTGGCC 778
Qy	1314	AGATTGTGCTCAACAGATCAAAATACTTAAGTTTGTGATGATGAAGCTGATCGATTTGGATA 1373
Db	779	GTATCTCCCTGTGACATCAAGTACCTCTGCTCTGATGAGGGCTGATTCGATCTTTGACA 838
Qy	1374	TGGTTTTTGGTCCAGAAATGAAGAATTAATTTTTCGCCAGGATGCCATCAAGAGAAC 1433
Db	839	TGGGTTTCGAGCTTCAAAATTTGCCGTATCGTTGAAGGTGAGACATCCCAAGTGAATG 898
Qy	1434	AGCGCAACCCCTTATGTCTAGTGCAACTTTTCCAGAGAAATTCAAAGSTTGGCTGCAG 1493
Db	899	ACCGCAGACTCTCATGTTCTCGGCCACTTTTCCCGCCGACATCCAAATCTCGCCCGCG 958
Qy	1494	AGTTTTTAAATCAAAATATCTGTTTGTGCTGTGACAAAGTGGGTGAGCATGTGAGAG 1553
Db	959	ATTTCCTTAAG--GACTCAATCTTCCGTGCGTTGGTGTGTTGTTTCCACTCTTGA 1015

QY 711 GAGGAGAAAGTGTATTACTCAAGGACCCAAAAGTGAACCTACATACCCCTCTCCACCTG 770
Db 311 GAAAGCGAGGGCTGGGACCGTAGGAGACGTGAAGTAAACCCCTTGAAAATGATGATTCGG 370
QY 771 AGGATGAGAGCTCCATCTTTGCACTATTATGAGACGAGCATAAACCTTGCAAAATTCGACA 830

Db	371	AACCAACACAGCTTTTACTGACGAGATATATACGGTTATTTATTTTGTATGCTTATGAAG	430
Qy	831	CTATCTTGTGGAAGTGTCTGACATGATGCAACCAAGCAATTCGACTTTTGAAGAAG	890
Db	431	ATATTCGATTGAGACCACTGGGGATAAATGTCCCTCCCTCTTAAACATTCGACAGA	490
Qy	891	CTAATCTGTGACACACTGAAATPACAATATGGTAAAGCTGTATATCTAAGCTTATC	950
Db	491	TGATCTTCGGAGGACATTGAATCTTAATAATCCGTATGACAAATATGTTAAGCCAAAC	550
Qy	951	CTGTGCAAAAATAACGATATTCCTATCATCTTCGACGACGAGATTGATGCTGTGTC	1011
Db	551	CTGTTCAGGCTATGCGATTCCGATATTCGATATTCCTTGAAGGGAGGATTTGATGGCTGTGCTC	610
Qy	1011	TAACCGGGCTGTGGGAAGACTGGGGCTTTTCTCTACCAATTTTGGCTCATATGATGATG	1071
Db	611	AGACGGGGTGGGGGAAGACAGCTGTTTTTGTTCCAATCATATGATGAATAAAGAAAG	670
Qy	1071	ATGGAATACTCCAG---TCGTTTAAAGAGTTCGACGAACCAAGATGATTTATTTGATG	1127
Db	671	ATCAGATGTATACAGAACCCCGTGGTTTACGAACAGCTACCTCTTGCAATATCTCT	730
Qy	1128	CACCACTCGAATTTGGTCAACCAATTTATTTGGAAGCCAGAAATTTCTTTTGGGA	1187
Db	731	CACCAACAGAGAGTTGGCAAGTCAAGTACATGATGAGGGCTAAAAAGTCTCTTATCCAA	790
Qy	1188	CTTGGTATAGAGCTGTGTTTATATATGAGGGGAACCCAGCTGGGACATTCATTTGACAA	1247
Db	791	CTGGTGTGAAGTGTGTTGCTATATGAGGAACCTTATTAACCAAGAGCTCCGGGAC	850
Qy	1248	TAGTACAAAGCTGTAATATATATATGTGTACTCTCTGGAAGCTGATGATATCATATGCA	1307
Db	851	TTGAGAGGGGAGTCGATATTTCTTGGGCAAGCCTGGTCGATTAATGATTTGCTCGAG	910
Qy	1308	AAGAAAAGATTGGTCTCAACAGATCCAAATCTAGTTTGGATGAAGCTATGCCATGT	1367
Db	911	GAGCTAGAGCTCTCAATGCATGATGATTAATTTTATGCTCTTGATGAGGCCATATGAATGC	970
Qy	1368	TGGATATGGGTTTGGTCCAGAAATGGAAGATTATTTCTTGCCCAAGATGCCATCAA	1427
Db	971	TTGACATGGGTTTGAACCAAAATTTGAAGATTTGTGCAACAAATGACATGCTCTCAC	1030
Qy	1428	AGGAACAGCCGCAAACTTATGTCTCAGTGCACATTTTCCAGAGAAATTAAGGTGG	1487
Db	1031	GTGAGTTTACAGACACTGTGTTATGTGTACATTTTCCAAAGAAATTCAGAGACTCG	1090
Qy	1488	CTGACAGATTTTAAAGTCAATATATCTGTTTGTGCTGTGACCAAGTGGGTGAGCAT	1547
Db	1091	CAGCTGACTTCTG---CAATATATATTTTTTGGCTGTGGTATGAGTGGGTTCAAGTA	1147
Qy	1548	GTAAGATGTTCAAGCAACGCTTCTCCAGTTGGCCAGTTCTCAAAAAGAAAAGCTCG	1607
Db	1148	CCGATTAATTTGTCCAAAGGGTTAGTTGTCTCTGACTGTGACAAAAGAGTACATCTCA	1207
Qy	1608	TTGAATTTCTGCAAAACATAGGGGATGAAGA-----ACTATGG	1646
Db	1208	TGGACCTGCTTCAAGCTCAGAGAGAAATGGCATCCAGCAAGCAAGCCCTGACTTGA	1267
Qy	1647	TCCTTGTGAACTAAGAAAAAAGACAGATTTTACTGCACTTTTCTTGTCAAGAAAAA	1706
Db	1268	TTTTTGTGGAGACAAAGAGAGAGCTGACTTTTGGAAAATGTGTTGTGATCAATGGGT	1327
Qy	1707	TATCAACTACAGATTCATGTGTATGCGGGAACAGAGAGCGGGAGCAAGCTCTTGGAG	1766
Db	1328	TTCCAGCAACCTCATTCACGGGTACGAACACAGCAGGAAAAGAGTGGCATTTGAAG	1387
Qy	1767	ATTTTCGCTTGGAAAATGCCCCAGTTCTTGTGTCTACTTCAATAGCTCCAGAGGGCTGG	1826
Db	1388	CCCTCAAGAGTGGGAGAACCCGATTTTGTGTGCATGTATGACAGACACGTGGCTTG	1447
Qy	1827	ATATGAAAATGTGCAACATGTTATCAATTTTATCTTCTTACCATGATGAATATG	1886

Db	1448	ACATGCCCCACGTGGCTCAGTGTCACTTCGATCTGCCAATGACATTGATGACTATG	1507
Qy	1887	TTTATGGAAATTGGGCGTACTGTGCTTTGGGGAATACTGGCAGAGCAATTCCTTTTGG	1946
Db	1508	TCCACCGTATTGGACGACGACGTGTGGCAATCAGGACTAGCACTGCTTCTTCA	1567
Qy	1947	ATCTTGATCGATTAACCAATTAGCAACGCTCTAGTAAAGTATTGACGATGCTCAAC	2006
Db	1568	A---CGATGGCAACACTCTCACTGCGCCAGACCGCTCGCTGAGCTGATGCAAGAACTAAC	1624
Qy	2007	AGGATGTTCTCTGCATGCTT	2025
Db	1625	AAGAAATCCTCTGAGTGGCT	1643

Search completed: July 27, 2004, 13:54:46
 Job time : 2093 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 05:15:24 ; Search time 166 Seconds
(without alignments)
7435.005 Million cell updates/sec

Title: US-09-714-865b-1
Perfect score: 2224
Sequence: 1 actgaagcaccacgsggsgg.....aagctctgtgttgatgca 2224

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	22.2	504	4	US-09-621-976-2923 Sequence 2923, Ap
2	388.8	17.5	420	4	US-09-833-381-1475 Sequence 1475, Ap
3	386.4	17.4	2451	4	US-09-976-584-786 Sequence 786, App
4	383.2	17.2	2319	3	US-09-058-489-90 Sequence 90, Appl
5	383.2	17.2	4416	3	US-09-058-489-17 Sequence 17, Appl
6	369.4	16.6	3408	3	US-09-058-489-14 Sequence 14, Appl
7	369.4	16.6	5322	3	US-09-183-706-42 Sequence 42, Appl
8	158.2	7.1	2365	3	US-09-567-995-42 Sequence 42, Appl
9	158.2	7.1	2365	4	US-09-328-352-509 Sequence 509, App
10	129.4	5.8	1191	4	US-09-976-584-213 Sequence 213, App
11	129	5.8	3760	4	US-09-543-681A-1401 Sequence 1401, App
12	127.6	5.7	1410	4	US-09-543-681A-1401 Sequence 1, Appl
13	125.6	5.6	1830121	4	US-09-557-884-1 Sequence 1, Appl
14	125.6	5.6	1830121	4	US-09-643-990A-1 Sequence 894, App
15	123.6	5.6	1254	5	PCT-US96-05320A-894 Sequence 894, App
16	123.6	5.6	1830121	4	US-09-557-884-1 Sequence 1, Appl
17	123.6	5.6	1830121	4	US-09-643-990A-1 Sequence 1, Appl
18	116.8	5.3	3230	4	US-08-961-527-203 Sequence 203, App
19	115.2	5.2	3347	4	US-09-702-705-318 Sequence 318, App
20	115.2	5.2	3347	4	US-09-702-705-318 Sequence 318, App
21	115.2	5.2	3347	4	US-09-614-124B-318 Sequence 318, App
22	115.2	5.2	3347	4	US-09-671-325-318 Sequence 318, App
23	115.2	5.2	3347	4	US-09-589-184-318 Sequence 318, Appl
24	113.2	5.1	3825	3	US-09-208-742-3 Sequence 3, Appl
25	113.4	5.1	1317	4	US-09-543-681A-195 Sequence 195, App
26	110.4	5.0	45613	4	US-09-596-002-22 Sequence 22, Appl
27	110.4	4.9	1941	4	US-09-328-352-1891 Sequence 1891, Ap

C	28	109.6	4.9	573	4	US-09-702-705-1384	Sequence 1384, Ap
C	29	109.6	4.9	573	4	US-09-736-457-1384	Sequence 1384, Ap
C	30	109.6	4.9	573	4	US-09-614-124B-1384	Sequence 1384, Ap
C	31	109.6	4.9	573	4	US-09-671-325-1384	Sequence 1384, Ap
C	32	106	4.8	7363	4	US-08-986-171E-19	Sequence 19, Appl
C	33	104.4	4.7	905	4	US-09-976-594-788	Sequence 788, App
C	34	103.2	4.6	1587	4	US-09-134-001C-1425	Sequence 1425, Ap
C	35	92.2	4.1	1245	3	US-09-318-443-1	Sequence 1, Appl
C	36	92	4.1	1365	4	US-09-543-681A-3545	Sequence 3545, App
C	37	91.2	4.1	1368	4	US-09-134-001C-816	Sequence 29, Appl
C	38	91.2	4.1	66985	4	US-09-596-002-29	Sequence 29, Appl
C	39	91	4.1	267	4	US-09-313-284A-3206	Sequence 3206, App
C	40	90.8	4.1	490	4	US-09-222-575-126	Sequence 126, App
C	41	90.8	4.1	490	4	US-09-389-681-126	Sequence 126, App
C	42	90.8	4.1	490	4	US-09-620-405B-126	Sequence 126, App
C	43	90.8	4.1	490	4	US-09-339-318-126	Sequence 126, App
C	44	90.8	4.1	490	4	US-09-433-826B-126	Sequence 126, App
C	45	90.8	4.1	490	4	US-09-604-287A-126	Sequence 126, App

ALIGNMENTS

RESULT 1									
US-09-621-976-2923									
Sequence 2923, Application US/09621976									
Patent No. 6639063									
GENERAL INFORMATION:									
APPLICANT: Dumas Milne Edwards, J.B.									
APPLICANT: Giordano, J.Y.									
TITLE OF INVENTION: ESTs and Encoded Human Proteins.									
FILE REFERENCE: GENSET.054PR2									
CURRENT APPLICATION NUMBER: US/09/621.976									
CURRENT FILING DATE: 2000-07-21									
NUMBER OF SEQ ID NOS: 19335									
SOFTWARE: Patent.pm									
SEQ ID NO 2923									
LENGTH: 504									
TYPE: DNA									
ORGANISM: Homo sapiens									
FEATURE:									
NAME/KEY: CDS									
LOCATION: 224..472									
US-09-621-976-2923									
Query Match									
Best Local Similarity 99.8%; Pred. No. 8.9e-134;									
Matches 504; Conservative 0; Mismatches 0; Indels 1; Gaps 1;									
QY	775	TGAGACCTCCATCTTGGACATTATCAGACGACATTAACCTTGACAAATACGACATAT	834						
DB	1	TGAGACCTCCATCTTGGACATTATCAGACGACATTAACCTTGACAAATACGACATAT	60						
QY	835	TCCTTGGAGAGTGTGACATGATGACACACCCAGCAATTCGACTTTGAAGAAGCTTA	894						
DB	61	TCCTTGGAGAGTGTGACATGATGACACACCCAGCAATTCGACTTTGAAGAAGCTTA	120						
QY	895	TCCTTGGAGAGTGTGACATGATGACACACCCAGCAATTCGACTTTGAAGAAGCTTA	954						
DB	121	TCCTTGGAGAGTGTGACATGATGACACACCCAGCAATTCGACTTTGAAGAAGCTTA	180						
QY	955	GCAAAATACAGTATTCCTATCATCTTGCAGACAGATTTGATGCTTGCTCAAC	1014						
DB	181	GCAAAATACAGTATTCCTATCATCTTGCAGACAGATTTGATGCTTGCTCAAC	240						
QY	1015	AGGCTCTGGAGAACTGCGCTTTCTCTACCAATTTGGCTCATATGATGATGATG	1074						
DB	241	AGGCTCTGGAGAACTGCGCTTTCTCTACCAATTTGGCTCATATGATGATGATG	300						
QY	1075	AATACCTGACGATCTTTTAAAGATTGGAGAACCCAGATGATATTGTAGACCAAC	1134						
DB	301	AATACCTGACGATCTTTTAAAGATTGGAGAACCCAGATGATATTGTAGACCAAC	360						

QY 1135 TCAGAAATTGGTCAACCAAGATTATTGGAAAGCCAGAAATTTCTTTGGGACTGTGT 1194
 DB 361 TCAGAAATTGGTCAACCAAGATTATTGGAAAGCCAGAAATTTCTTTGGGACTGTGT 420
 QY 1195 AAGAGCTGTGTATATATATGGGGGAACCCAGCTGGGACATTCATTGCAAAATAGTACA 1254
 DB 421 AAGAGCTGTGTATATATATGGGGGAACCCAGCTGGGACATTCATTGCAAAATAGTACA 479
 QY 1255 AGGCTGTATATATATATATGCTACT 1279
 DB 480 AGGCTGTATATATATATATGCTACT 504

RESULT 2

US-09-833-381-1475
 ; Sequence 1475, Application US/09833381
 ; Patent No. 6672186
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 5800-119
 ; CURRENT APPLICATION NUMBER: us/09/833,381
 ; PRIOR FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 09/516,448
 ; NUMBER OF SEQ ID NOS: 2050
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1475
 ; LENGTH: 420
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-833-381-1475

Query Match 17.5%; Score 389.8; DB 4; Length 420;
 Best Local Similarity 99.5%; Pred. No. 9,9e-104;
 Matches 391; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 818 GAAATATACACACTTCTTCTGTGGAAGTGTGACATGATGACACCAACCAATTCG 877
 DB 28 GTCAATATACACACTTCTTCTGTGGAAGTGTGACATGATGACACCAACCAATTCG 87
 QY 878 ACTTTGAAGAGCTATCTCTGTGACACCTGAATTAACAATTCGTAAGCTGTAT 937
 DB 88 ACTTTGAAGAGCTATCTCTGTGACACCTGAATTAACAATTCGTAAGCTGTAT 147
 QY 938 ACTTAAGCTATCTCTGTGCAAAAATACAGTATTCCTATCATCTTGCAGAGAGATTG 997
 DB 148 ACTTAAGCTATCTCTGTGCAAAAATACAGTATTCCTATCATCTTGCAGAGAGATTG 207
 QY 998 ATGGCTTGTCTCAACAGAGTCTGGAGAGACTGGGGCTTTCTCTCAATTTGGCT 1057
 DB 208 ATGGCTTGTCTCAACAGAGTCTGGAGAGACTGGGGCTTTCTCTCAATTTGGCT 267
 QY 1058 CATATGATCATGATGGAATTAACCTGCACTCGTTTAAAGATTGCAAGAACAGAGTGT 1117
 DB 268 CATATGATCATGATGGAATTAACCTGCACTCGTTTAAAGATTGCAAGAACAGAGTGT 327
 QY 1118 ATTATGTGACCAACCTGCAAAATTTGGTCAACCAAGATTATTGGAAGCCAGAAATTT 1177
 DB 328 ATTATGTGACCAACCTGCAAAATTTGGTCAACCAAGATTATTGGAAGCCAGAAATTT 387
 QY 1178 TCTTTGGGACTTGTGAAGAGCTGTGTATA 1210
 DB 388 TCTTTGGGACTTGTGAAGAGCTGTGTATA 420

RESULT 3

US-09-976-594-786
 ; Sequence 786, Application US/0976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael

APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 786
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 035282CB1
 ; US-09-976-594-786

Query Match 17.4%; Score 386.4; DB 4; Length 2451;
 Best Local Similarity 58.7%; Pred. No. 2,7e-102;
 Matches 763; Conservative 0; Mismatches 501; Indels 36; Gaps 4;

QY 801 AGACAGCATTAACCTTGACAAATAGACACTATTCCTTGTGGAAGTGTGACATGATG 860
 DB 503 ACACGGGATTAACCTTGAGAAATATGATATACAGTAGAGCAACCGGAGTACT 562
 QY 861 CACCACAGCAATCTGACTTTTGAAGAGCTATCTCTGACACCTGATTAACACA 920
 DB 563 GTCTCCACATATTGAATTTTACGATATTCGATGAGAGAAATATATATATGAGGAGACA 622
 QY 921 TTGCTAAAGCTGTATATCTAAGCTTACTCTCTGTGCAAAAATACAGTATTCATATAC 980
 DB 623 TTGAATTAACCTGCTATATCTGCTCTCTCTGCAAAAATACAGTATTCATATATTA 682
 QY 981 TTGAGAGACAGATTTATATGCTGTGCTCAACAGGCTTGGGAAGACTGGGCTTTTC 1040
 DB 683 AGGAAAAAGAGATTAATGCTTGTGCCCAACAGAGATCGGAAAACTGCAGCAATTC 742
 QY 1041 TCTTACCATTGCTCA-----TATGATCATGATG 1073
 DB 743 TTTTACCATATGAGTCAATATATACAGATGCTCAGAGAGAGCTTTGAAGCTGTGA 802
 QY 1074 GAATTAATGCAAGTCTGTTTAAAGAGTTGCAAGAACCAAGTATATTATGACCAA 1133
 DB 803 AGGTAATGGAAGTATGAGGCGCGCAACATATCTGTTTGAAGCTTGA 862
 QY 1134 CTCGAGATTGCTCAACAGATTTATTTGGAGCCAGAAATTTCTTTGGGACTGTG 1193
 DB 863 CAGAGATTGCTGTACAGATCTATGAGAGACCAAAATTTCTTACCGATTTAGAG 922
 QY 1194 TAAAGCTGTTTATATATGAGGGAACCCAGCTGGACATTCATATGACAAATAGTAC 1253
 DB 923 TTGCTCTGTGTATATATGAGTGTGCTGATTTGTGACAGATTCGGGACTTGAAC 982
 QY 1254 AAGGCTGTATATATATGATGCTACCTGGAAGACTGATGATATCATAGCCAAAGAA 1313
 DB 983 GTGATGCTCACTGTTATGATGCTACCTGCAAGAGCTATGATGATGAGAAAGGGA 1042
 QY 1314 AGATTGCTCAACAGATCAATATCTAGTTTGGATGAAGCTGATGATGATGATA 1373
 DB 1043 AGATTGATATGATCTGCAAGTACTTGTGATGATGATGATGATGATGATGATGATA 1102
 QY 1374 TGGGTTTGGTCCAGAAATGAAGAGTATTTCTTCCAGAAATGCCATTAAGGAAC 1433
 DB 1103 TGGGATTTGAACCTGATGATGCTGATGATGATGATGATGATGATGATGATGATGATA 1162
 QY 1434 AGGCAACACCTTATGCTCAGTCACTTCCAGAGAAATTAAGAGTTGGCTGAG 1493
 DB 1163 TTGCTACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 1222
 QY 1494 AGTTTAAAGTCAATTTATCTGTTTGTGCTGTTGACCAAGTGGGTGAGATGAG 1553
 DB 1223 ACTTTTGGATGAA--TATATCTTTTGGCTGTGAGGAGAGTATGATGATGATGATGATA 1279

QY 1554 ATGTTGAGCAGACCGCTTCTCCAGTTGGCCAGTCTCAAAAAGAGAAAGCTGTTGAA 1613
Db 1280 ACATCACACAGAAAGAGTGGTGGTGAAGACTTAAGATAAGCGTCATTTCTACTGACA 1339
QY 1614 TTCTGCGAAACATAGG---GATGAAAGAACTATGCTCTTTGTAAGTAAAGAAAAG 1670
Db 1340 TTTTAGGTGCAACAGGAGAGTACTTACTTACTTACTTGTGTGGAGACCAAAAAGGGAG 1399
QY 1671 CAGATTTTACTGCACTTTTCTTGTGCGAAGAAAATATCACTACATGATTCATGTG 1730
Db 1400 CAGATTTCCCTGGAGATTTCTTATACATGAAAGATGCTGTACTATATTCATGAG 1459
QY 1731 ATGGGGAACAGAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGGAAAGTGGCCAG 1790
Db 1460 ACCGGTCAACAGAGATCGAGAGAGGCCCTTCAACAGTTTGGCTCAGGAAAAGCCCA 1519
QY 1791 TTCTTGTGCTACTTCACTAGTGTGCAAGAGGCTGATATGAAAATGTGCAACATGTA 1850
Db 1520 TTCTAGTGTCTACAGCTGTGGCAGACAGAGACTAGACATTTCAATGTGAGCATGTA 1579
QY 1851 TCAATTTGATCTCTCTTACATGATGATATGATGATGCAATGGAATGGCGTACTGTGC 1910
Db 1580 TCAATTTGATTTGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1639
QY 1911 GTTGTGGGAATACTGCGAGAGCAATTTCTTTTGTGATCTTGATCGATTAACCATTTAG 1970
Db 1640 GTGTAGAAACCTGGGCTTGGCACCCTCATTTCTTTA---TGAAAAAATATGATATTA 1696
QY 1971 CACAGCTCTAGTAAAGTATGACAGATCTCAACAGATGTTCTCTGATGTTGGAAG 2030
Db 1697 CAAAGGATTTGTTGATCTTCTTGTAGAGCTTAAACAGAGAGGCTTCTTGTGTTGAAA 1756
QY 2031 AAATGGCTTTAGACATACATTCCTGGCTTCAGTGTAG 2070
Db 1757 ATATGGCTTATGAACACCACTACAGAGGTGGCAGTGTG 1796

RESULT 4

US-09-058-489-90
; Sequence 90. Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lam, Bruce
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-089A
; CURRENT APPLICATION NUMBER: US/09/058, 489
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 90
; LENGTH: 2319
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-90

Query Match 17.2% Score 383.2; DB 3; Length 2319;

Best Local Similarity 58.5%; Pred.No.2.2e-101;

Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

QY 801 AGACAGCATTAACCTGACAAATAGCACTATTTCTTGGAAGTGTGCAATGATG 860
Db 529 ACAGGGGATTAACCTTGAGAAATATGATGATATACAGTAGAGGCAACGGCAGTACT 588
QY 861 CACACAGCAATTTCTGCTTTTGAAGAGTATCTGTGACAGACGTAATTAACACA 920
Db 589 GTCTCCACATATTGAGAAATTTAGCGATTTGACATGGGAGAAATTAATCATGGGGACA 648

QY 921 TTGCTAAAGCTGTTATCTAAGCTTACTCTGTGCAAAAATACAGTATTCATATAC 980
Db 649 TTGAACCTTACTCGCTATATCTGTCTCTACTCCAGTGCAGAAAAACATGCCATTCATATTA 708
QY 981 TTGACAGCAGATTTGATGCTGTGTCTCAAAAGAGCTCTGGAAAGACTGGGCTTTTC 1040
Db 709 AGGGAAGAAAGACTAGTGGCTGTGTGCTGCAAAAGAGATCTGGAGAAACCTGCACATTTTC 768
QY 1041 TCCATCAATTTGGCTCA-----TATGATGATGATG 1073
Db 769 TTTTACCATCTAGTCTAGATATATATACAGATGCTCCAGAGAAAGCTTTGAAGCTGTGA 828
QY 1074 GAATATCTGCCAGTCTTTTAAAGATGTCAGAAACCAAGTGTATTTATGTAGACCAA 1133
Db 829 AGAAAAATGAAAGGATATGGGCGCCGCAAAACATATCCATATCTTGGTTTACCCCCCA 888
QY 1134 CTGAGATTTGTGCAACAGATTTATTTGAAAGCCAGAAAATTTCTTTGGGACTGTG 1193
Db 889 CAAGGAATTTGCTGTACAGATCTATGAGAGCCAGAAAATTTTCTTACCATCTAGAG 948
QY 1194 TAAGAGCTGTTGTTATATATGAGGAGAACCCAGCTGGGACATTCATTCACAAATAGTAC 1253
Db 949 TTGCTCTTGTGATGTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 1008
QY 1254 AAGCTGTATATATATATATGCTACTCTCTGGAAGACTGATGATATCATAGCAAGAAA 1313
Db 1009 GTGATGCTCCTGTTGATGACCACTCCAGAGAGCTGATGATGATGATGATGATGATGAT 1068
QY 1314 AGATGCTCTCAACAGATCAAACTAGTTTGTGATGAGATGATGATGATGATGATGATGAT 1373
Db 1069 AATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1128
QY 1374 TGGGTTTGTCTCAGAAATGAAAGATTAATTTCTTCCAGAAATGCTCATPAAAGAAC 1433
Db 1129 TGGGATTTGAACCTCAGATAGCTGTATGATGAAACAGATATATGCTCACCAGAGGCG 1188
QY 1434 AGCGCAACCTTATGTTGCTAGTGTGCACTTTCCAGAGAAATTCAGAGTTGGCTGAG 1493
Db 1189 TTGCTCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1248
QY 1494 AGTTTAAAGTCAAAATATCTGTTTGTGCTGTGCAAGATGGGTGAGCATGTAGAG 1553
Db 1249 ACTTTTGGATGAA---TATATCTTTTGGCTGTGAGAGATGAGTCTACCTGTGAGA 1305
QY 1554 ATGTTGACGACAGCTTCTCAAGTGGCCAGTCTCAAAAAGAGAAAGCTCTGTTGAA 1613
Db 1306 ACATCACACAGAAAGTATGTTGGGTGAAAGCTTATGATTAACGGTCAATTTCTACTGACA 1365
QY 1614 TTCTGCGAAACATAGG---GATGAAAGAACTATGCTTTGTAAGTAAAGAAAAG 1670
Db 1366 TTTTAGGTGCAACAGGAGAGTATTTCACTTATGATGATGATGATGATGATGATGATGAT 1425
QY 1671 CAGATTTTACTGCACTTTTCTTGTCAAGAAAATATCACTACATGATCCATGTGTG 1730
Db 1426 CAGATTTCCCTGGAGATTTCTTATACATGAAGATGCTGTGATCTGATATTCATGAG 1485
QY 1731 ATGGGGAACAGAGAGAGGAGGAGCAAGCTCTGAGATTTTGGCTTTGAAAAGTCCAG 1790
Db 1486 ACCGGTCAACAGAGATGAGAGAGGCCCTTCAACAGTTTGGCTCAGGAAAAGCCCA 1545
QY 1791 TTCTTGTGCTACTTCACTAGTGTGCAAGAGGCTGATATGAAAATGTGCAACATGTA 1850
Db 1546 TTCTAGTGTCTACAGTGTGCGAGCAGAGACTAGATTTCAATGATGTGAGACATGTA 1605
QY 1851 TCAATTTGATCTTCTTCTTCACTTATGATGATGATGATGATGATGATGATGATGATGAT 1910
Db 1606 TCAATTTGATTTGCAAGTATATGAAAGATGATGATGATGATGATGATGATGATGATGAT 1665
QY 1911 GTTGTGGGAATCTGCGAGAGCAATTTCTTTTGTGATCTTGAATCGATTAACATTTAG 1970
Db 1666 GTGTAGAAACCTGGGCTGTGCACTCATCTTTTA---TGAAAAAATATGATATTA 1722
QY 1971 CACAGCTCTAGTAAAGTATGACAGATGCTCAACAGATGTCTCTGATGTTGGAG 2030

Db 1723 CAAGAGATTGGTGGATCTTCTGTAGAGCTAAACAGAGAGCTTCTGTGGTGA 1782
Qy 2031 AAATGCTTTAGTACATATCTTCGGCTTCAGTGTAG 2070
Db 1783 ATATGGCTTATGAACACCACTACAGAGGTGGCAGTGTGG 1822

RESULT 5
US-09-058-489-17

/ Sequence 17, Application US/09058489
/ Patent No. 610386
/ GENERAL INFORMATION:
/ APPLICANT: Whitehead Institute for Biomedical Research
/ APPLICANT: Lahn, Bruce
/ APPLICANT: Page, David
/ TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of
/ TITLE OF INVENTION: the Y Chromosome
/ FILE REFERENCE: WHI97-08PA
/ CURRENT APPLICATION NUMBER: US/09/058,489
/ EARLIER FILING DATE: 1998-04-10
/ EARLIER APPLICATION NUMBER: 60/041,877
/ EARLIER FILING DATE: 1997-04-11
/ NUMBER OF SEQ ID NOS: 91
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 17
/ LENGTH: 4416
/ TYPE: DNA
/ ORGANISM: Human
US-09-058-489-17

Query Match 17.2%; Score 383.2; DB 3; Length 4416;
Best Local Similarity 58.5%; Pred. No. 3.2e-101;
Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

Qy 801 AGACAGGCAATTAACCTTGCACAAATAGACATATTTGTGGAAGTGTGACATGATG 860
Db 529 ACAGGGGATTAACCTTGAAGAAATATGATGATATACATAGACACCGCACTAAT 588
Qy 861 CACCACGACAAATCTGACTTTGAAAGACTATCTTGTGACACTGAATTAACA 920
Db 589 GTCCCTCACAATATGGAATTTTGAAGATTTGACATGGAAGAAATTAATCATGGGAA 648
Qy 921 TTGCTAAGCTGTTTACTAAGCTTACTCTGCTGGAAGAAATATACATTTCTATCATAC 980
Db 649 TTGAATCTTACTGCTATCTCTGCTCTACTCTGCTGCAAGCAAAACATGCCATTTCTATTA 708
Qy 981 TTGCAGACGAGATTTGATGCTGTGCTCAACAGAGCTTGGGAAGACTGGGCTTTTC 1040
Db 709 AGGAAAAAGAGCTAGTGGCTTGTGCCCCAAGAGATCTGGAAAACTGCAGATTTTC 768
Qy 1041 TCCTACCAATTTGGCTCA-----TATGATGATGATG 1073
Db 769 TTTTACCCACTGAGTCAAGATATATACAGATGTCAGAGAGACTTTGAAGGCTGTGA 828
Qy 1074 GAATTAATGCAAGCTGTTTAAAGAGTTGCAGAAACAGAGTGTATTTATGTAGACCAA 1133
Db 829 AGGAATATGAAAGTATGGGCGCCGCAAAATATCAATATCTTGGTTTAAAGCCAA 888
Qy 1134 CTGAGAAATTTGTCACACGATTTATTTGGAAGCCAGAAATTTCTTTTGGAGCTTGTG 1193
Db 889 CAAGAAATTTGCTGTACAGATCTATGAGAAAGCCAGAAATTTTCCATCCGATTTAG 948
Qy 1194 TAAGAGCTGTTTATATATGAGGGAACCCAGCTGGGACATTCATTTGACCAAAATAGTAC 1253
Db 949 TTGCTCTTGTATGATTTATGTTGTGTGTGCTGATATTTGTCAGAGATTTGGGACTTGAAC 1008
Qy 1254 AAGGCTGTATATTTATGTTGTCTCTCTGGAAGACTAGATATCTAGGCAAGAAA 1313
Db 1009 GTGAGTGCACACTTTAGTACCACTCCAGAGCTGTGATGATATGATGAAAGAGAA 1068
Qy 1314 AGATTGGTCTCAAAACGATCAATATCTAGTTTGGATGAGAGCTGATGCGATGTTGATA 1373

Db 1069 AGATTGATTAAGCTTCTGCAAGTACTTATGTTGATGAAGCTGATAGGATGCTGATA 1128
Qy 1374 TGGTTTTTGTCCGAAGATGAAGATTAATTTCTTCCAGAGATGCCATCAAGGAAC 1433
Db 1129 TGGATTTGAACCCACATAGCTGTATATGTAACAAGATATATGCCACCAAGGGCG 1188
Qy 1434 AGCCCAAAACCTTATGTCAGTCACTTTTCAGAGAGAAATTCAGAGGTTGGCTGAG 1493
Db 1189 TTGTCACACCATATATTTAGTGTCTCTTTCTTAAGAAATATACAGATCTGCTGTG 1248
Qy 1494 AGTTTAACTCAATATATCTGTTTCTGCTGTGAGACAAGTGGTGGACATGTAGAG 1553
Db 1249 ACTTTTGGATGAA---TATATCTTTTGGCTGTAGGACAGATAGGCTTACCTGTAGA 1305
Qy 1554 ATGTTGACGAGACCGTCTCCAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAA 1613
Db 1306 ACATCAACAGAAAGTATTTGGTGGAAACCTTATGTAACGTCTATTTCTACTGACA 1365
Qy 1614 TTCTGCAAACTATGGG---GATGAAGAACTATGCTTTTGTGAAACTAGAAAAAG 1670
Db 1366 TTTTAGCTGCAACAGGAGTGTATCTTACTTATGTTGTGGAGACCAAAAAGGGAG 1425
Qy 1671 CAGATTTTACTGCACTTTCTTTGCAAGAAATATCACTAGATAGTCAATGCTG 1730
Db 1426 CAGATTTCTGAGAGATTTCTTATACATGAGATATGCTTGTACTATGATATTCATGAG 1485
Qy 1731 ATCGGAAACAGAGAGACGGGAGCAAGCTTTGAGATTTTCGCTTTGGAAGTGGCCAG 1790
Db 1486 ACCGTACAGAGAGATCGAGAGAGGCCCTTCAACAGTTTGGCTCAGGAAAAAGCCCA 1545
Qy 1791 TTCTTTGCTACTTCAATGCTGCGAGAGGCTGATATTTGAAATTTGCAACATGTA 1850
Db 1546 TTCTAGTGGTATACGCTGTGAGACAGAGACTTGAATTAATGAGACATGTA 1605
Qy 1851 TCAATTTGATCTTCTTACCATTTGATGAATATGTTTCAATTTGGGCGTACTGTC 1910
Db 1606 TCAATTTGATTTGCCAAGATATTTGAAGATATGATGATGATGATGATGATGATGAT 1665
Qy 1911 GTTGTGGAATCTGCGACAGCAATTTCTTTTGAATCTTGAATGCAATPACATTTAG 1970
Db 1666 GTGAGGAAACCTGGGCTTGCACCTCAATCTTTAA---TGAAAAAATATGAATATTA 1722
Qy 1971 CACAGCTCTAGTAAAGTATGACAGATGCTCAACAGAGTGTCTGATGGTTGGAG 2030
Db 1723 CAAGAGATTTGATGATCTTCTGTGAAAGCTTAAACAAGAGTGCCTTCTTGTGGAAA 1782
Qy 2031 AAATGCTTTAGTACATATCTTCGCTTCAGTGTAG 2070
Db 1783 ATATGGCTTATGAACACCACTACAGAGGTGGCAGTGTGG 1822

RESULT 6

US-09-058-489-14
/ Sequence 14, Application US/09058489
/ Patent No. 610386
/ GENERAL INFORMATION:
/ APPLICANT: Whitehead Institute for Biomedical Research
/ APPLICANT: Lahn, Bruce
/ APPLICANT: Page, David
/ TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of
/ TITLE OF INVENTION: the Y Chromosome
/ FILE REFERENCE: WHI97-08PA
/ CURRENT APPLICATION NUMBER: US/09/058,489
/ EARLIER FILING DATE: 1998-04-10
/ EARLIER APPLICATION NUMBER: 60/041,877
/ EARLIER FILING DATE: 1997-04-11
/ NUMBER OF SEQ ID NOS: 91
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 14
/ LENGTH: 3408
/ TYPE: DNA
/ ORGANISM: Human
US-09-058-489-14

Query Match 16.6%; Score 369.4; DB 3; Length 3408;
 Best Local Similarity 57.6%; Pred. No. 2,9e-97;
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

786 TCTTTGACATTTACAGACAGGATTAACCTTGACAAATTCGACACTATTTCTTGAGAG 845
 1305 TCTTTTCTGGAGGCAACACTGGGATTTATTTAGAAATACGATGACATTTCCAGTTGAG 1364
 846 TGTCTGGACATGATGACACACACGCAATTCGACTTTTGAAGAAGTAATCTCTGACAG 905
 1365 CAACAGGCAACACTGTCCTCCACATTTGAAAGTTTCACTGATGTTGAGATGGAGAAA 1424
 906 CACTGAATTAACACATTGCTTAAGCTGTTATTAAGCTTACTCTGTCGCAAAAATACA 965
 1425 TTATCATGGGAAACATTGAGCTTACTGCTTAACCTGCCCACTCCAGTGCAGAAAGCATG 1484
 966 GTATTCCTATCATCTTGGACAGAGAGATTTGATGCTTGCTCAACAGGGCTGAGG 1025
 1485 CTATTCCTATTTATCAAGAGAAAAGAGACTTGATGCTTGCTGCCAACAAGGGCTGAG 1544
 1026 AGACTGGCGCTTTCTCTCAACCAATTTTGCTCATATGATGATGATG----- 1074
 1545 AAACCTGACGATTTCTGTCGCCCATCTTGAGTCAGATTTATTCAGATGCTCCAGGCGAG 1604
 1075 -----AATACTGCCAGTGGTTTAAAGAGTTGCGAGAACGACAGTGA 1118
 1605 CTTTGAGGGCCATGAGAAATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1664
 1119 TTATTTAGACACCACTCGAGATTTGCTCAACAGATTTATTTTGAAGCCAGAAATTTT 1178
 1665 TGTATTTAGACACCAACAGAGAGATTTGACAGATCTGAGAAAGCCAGAAATTTT 1724
 1179 CTTTGGAGCTTGTGTAGAGCTGTTTATATATGAGGGAACCCAGCTGGACATTCGA 1238
 1725 CATCCCATCTAGAGTTGCTGCTGCGGTGTTATGTTGTCGCGATTTGGTTCAGACA 1784
 1239 TTGCAAAATAGTCAAGGCTGTATATATATATGCTACTCTCCGGAAGACGATGATA 1298
 1785 TTCCAGACTTGGAAACGTGATGCCATTTGTTAGAGCCATCCGAGAGCGTCTAGTGA 1844
 1299 TCATAGGCAAGAAAGATTTGCTCAACAGATCAAAATCTTATGTTTGAATGAGCTG 1358
 1845 TGATGGAAGAGAGAAAGATTTGATTTAGCTTTTCAAAATCTTGTGTTGATGAGAGCTG 1904
 1359 ATCCGATGTTGATATGAGGTTTGTGTCAGAAATGAAAGATTAATTTTGGCCAGAA 1418
 1905 ATCGAGTTGATATGAGGTTTGTGAGCTTCAGATTCGTAGATGTTGGAACAGATCTGA 1964
 1419 TGCCATCAAGAAACAGGCGCAACCCCTTATGTTCACTGCACTTTTCCAGAGAAATTC 1478
 1965 TGCTCCCAAGAGGTCGCGCACACATGATGTTAGTTAGCTTATCTTCTTAAAGAAATAC 2024
 1479 AAAGTTGGCTGCAAGCTTTTAAAGTCAATTAATGTTGTTGCTGTTGGAACAATGG 1538
 2025 AGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2081
 1539 GTGAGCATGATAGATGTTGACAGACCGTTCCTCAAGTTGAGGCACTTCCAAAAAGAG 1598
 2082 GCTTACTCTGAAACATCAACAGAAAGATTTGGGTGGAAGATCAGACAAAGGT 2141
 1599 AAAAGCTGTTGAATTTGCGAAACATAG--GATGAAAGAACTATGCTCTTTGTTG 1655
 2142 CATTTCTGCTTGAACCTCTTAATATGCAACAGGCAAGGATTCAGTCACTTAATGTTTGG 2201
 1656 AAACCTAAGAAAAACAGATTTTACTGCACTTTTCTTCTCAAGAAAAAATATCACTA 1715
 2202 AGACCAAAAAAGGTGCAATTTCTGAGAGATTTCTTATACAGAAAGATACGCAATGA 2261
 1716 CAAGTATCATGATGATGAGGAAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGCT 1775
 2262 CCAGCATCATGAGACGTTCTCAGAGGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2321

1776 TTGGAAGTCCCACTTCTTGTGCTACTGATAGTCCAGAGGGCTGATTTGAAA 1835
 2322 CAGGAAAAAGCCATTTTATGAGCTACAGACGATGACGAAAGACTGACATTTCAA 2381
 1836 ATGTCAACATGTTATCAATTTTGTATCTTCTTCTTCACTTATGATGATATGTTACGAA 1895
 2382 ATGTAAACATGTTATCAATTTTGTATCTTCTTCTTCACTTATGATGATATGTTACGTA 2441
 1896 TTGGGCTGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1955
 2442 TTGCTGATGAGGAGAGTGTAGAAACCTTGCTGCTGCACTTCTTTTAA---CGGA 2498
 1956 CGGATACATTTAGACAGCCTCTAGTAAAGATTTAGACAGATGCTCAGAGATGTTTC 2015
 2499 GGAACATTAATTTCTAAGATTTGTTGATCTTCTGTTGAAGCTTAAACAGAGATGTC 2558
 2016 CTGATGTTGAGAGAAATGCTTCTTATGATACATATCTTGGCTTCAGTGTAG 2070
 2559 CGTCTGTTAGAAACATGCTTATGAAACCACTACAGAGGTAGACAGTGTG 2613

RESULT 7
 US-09-058-489-13
 ; Sequence 13, Application US/09058489
 ; Patent No. 6103886
 ; GENERAL INFORMATION:
 ; APPLICANT: Whitehead Institute for Biomedical Research
 ; APPLICANT: Lahn, Bruce
 ; APPLICANT: Page, David
 ; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
 ; FILE OF INVENTION: the Y Chromosome
 ; FILE REFERENCE: WH197-08PA
 ; CURRENT APPLICATION NUMBER: US/09/058,489
 ; EARLIER FILING DATE: 1998-04-10
 ; EARLIER APPLICATION NUMBER: 60/041,877
 ; EARLIER FILING DATE: 1997-04-11
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 5322
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-058-489-13

Query Match 16.6%; Score 369.4; DB 3; Length 5322;
 Best Local Similarity 57.6%; Pred. No. 3.8e-97;
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

786 TCTTTGACATTTACAGACAGGATTAACCTTGACAAATTCGACACTATTTCTTGAGAG 845
 1305 TCTTTTCTGGAGGCAACACTGGGATTTATTTAGAAATACGATGACATTTCCAGTTGAG 1364
 846 TGTCTGGACATGATGACACACGCAATTCGACTTTTGAAGAAGTAATCTCTGACAG 905
 1365 CAACAGGCAACACTGTCCTCCACATTTGAAAGTTTCACTGATGTTGAGATGGAGAAA 1424
 906 CACTGAATTAACACATTGCTTAAGCTGTTATTAAGCTTACTCTGTCGCAAAAATACA 965
 1425 TTATCATGGGAAACATTGAGCTTACTGCTTAACCTGCCCACTCCAGTGCAGAAAGCATG 1484
 966 GTATTCCTATCATCTTGGACAGAGAGATTTGATGCTTGCTCAACAGGGCTGAGG 1025
 1485 CTATTCCTATTTATCAAGAGAAAAGAGACTTGATGCTTGCTGCCAACAAGGGCTGAG 1544
 1026 AGACTGGCGCTTTCTCTCAACCAATTTTGCTCATATGATGATGATG----- 1074
 1545 AAACCTGACGATTTCTGTCGCCCATCTTGAGTCAGATTTATTCAGATGCTCCAGGCGAG 1604
 1075 -----AATACTGCCAGTGGTTTAAAGAGTTGCGAGAACGACAGTGA 1118
 1605 CTTTGAGGGCCATGAGAAATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1664
 1119 TTATTTAGACACCACTCGAGATTTGCTCAACAGATTTATTTTGAAGCCAGAAATTTT 1178

Db 1665 TGGTATTGACCAACGAGAGTGGACAGATCTACGAAAGAGCCAGAAAATTTT 1724
Qy 1179 CTTTGGGACTTGTGAAGCTGTGTATATATGAGGGAACCCAGCTGGACATTCA 1238
Db 1725 CATACCATCTAGAGTTCCTTCTGCTGTTATATGAGTGGCCGATATGGCTCAGAGA 1784
Qy 1239 TTGCACAATATGATCAAGGCTGTATATATATATATGAGTGGCCGATAT 1298
Db 1785 TTCCAGACTTGGACGAGATGACATTTGTAGACCACTCCAGAGCTCTAGTGATTA 1844
Qy 1299 TCATAGGCAAGAAAGATGAGTGTCTCAACAGATCAATCTAGTTTGGAGAACTG 1358
Db 1845 TGAATGAAAGAGAAATGAGATTAAGCTTTGCAATATCTGTTGATGATGAAGCTG 1904
Qy 1359 ATGCATGTTGATATGAGTGTGTTGTCAGAAATGAAAGATTAATTTCTGCGCAGAA 1418
Db 1905 ATCGATGTTGATATGAGGTTTGAAGCTCAGATTTGTATATATGATGAACAGATCTA 1964
Qy 1419 TGGCATTAAGGAACAGCGCCAAACCTTATGTTCACTGCACTTTTCCAGAGAAATTC 1478
Db 1965 TGCTCCAAAGGAGTGTCCGCAACATGATGTTTGTGTCTACTTTTCTTAAGAAATAC 2024
Qy 1479 AAGAGTGGCTGACAGATTTTAAAGCAATATCTGTTTGTGCTGTGACAGATG 1538
Db 2025 AGATGCTGCTCGATTTCTTAGATGA---TATATCTTGTGCTGTAGAGAGATG 2081
Qy 1539 GTGAGAGATGATGATGTTCCAGACAGCTTTCCAGTGGCCAGTTCTCAAAAAG 1598
Db 2082 GCTCTACCTCTGAAATATCACAGAAATGATTTGGGTGAAGATGACAAACGCT 2141
Qy 1599 AAAAGCTGTTGAAATCTCGAAACATAGS---GGATGAAGAACTATAGTCTTGTG 1655
Db 2142 CATTTGCTTGAACCTCTTAATGCAACAGGCAAGATTCAGTACCTTAGTGTG 2201
Qy 1656 AAATTAAGAAAGAGAGATTTTACTGCACTTTTCTTGTCAAGAAATATCACTA 1715
Db 2202 AGACCAAAAGGAGTGCAGATTTCTGAGAGATTTCTTATACATGAAGATACGATGA 2261
Qy 1716 CAAGTATCCATGAGATTCGGGAACAGAGAGCGGAGCAAGCTTTGAGATTTTGGT 1775
Db 2262 CAGACATCAATGAGACCTGTTCTCAGAGGATGAGAAAGAGCTTCAACAGTCCGCT 2321
Qy 1776 TTGGAAGTCCAGATTTCTGTTGCTACTTCACTAGTCCAGAGGCTGATTTGAAA 1835
Db 2322 CAGAAAGAGCCCAATTTAGTGGCTACAGAGTACAGAAAGAGACCTGGAATTTCA 2381
Qy 1836 ATGTGCAACATGTTATCAATTTGATCTTCTTCACTTGAATATGTTCAATGAA 1895
Db 2382 ATGTGAAACATGTTATCAATTTGATCTTCCAAATGATATGAAAGATATGACATGTA 2441
Qy 1896 TTGGGCGTACTGCTGTTGGGAATAGTGGCAGAGCAATTCCTTTTGTGATTTGAAT 1955
Db 2442 TTGGTGTAGGAGAGCTGTAGAGAACTTGGCTGCAACCTCATTTCTTAA---CGAGA 2498
Qy 1956 CGATTAACATTTAGACAGCTCTAGTAAAGATTTGACAGATGCTCAACAGATGTT 2015
Db 2499 GGAACATTAATATATTAAGATTTGTGTGATCTTCTTGTGAAGCTAAACAGAGGCT 2558
Qy 2016 CTGCATGTTGGAAGAAATTCCTTATGATACATTCCTGCTTCACTGAGTATG 2070
Db 2559 CGCTTGTGTAAAGACATGGCTTATGAACACCACTACAGGCTTACAGCTCTGG 2613

RESULT 8
US-09-183-706-42
; Sequence 42, Application US/09183706
; Patent No. 6245525
; GENERAL INFORMATION:
; APPLICANT: Martelange, Valérie
; APPLICANT: De Smet, Charles
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR

FILE REFERENCE: L0461/7054
CURRENT APPLICATION NUMBER: US/09/183,706
CURRENT FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: 09/122,989
EARLIER FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 43
SEQ ID NO 42
LENGTH: 2365
TYPE: DNA
ORGANISM: H. sapiens
NAME/KEY: CDS
LOCATION: (208)...(2151)
US-09-183-706-42

Query Match 7.1%; Score 158.2; DB 3; Length 2365;
Best Local Similarity 50.1%; Pred. No. 9.2e-36;
Matches 518; Conservative 0; Mismatches 488; Indels 27; Gaps 4;

Qy 915 ACAACATTGCTAAAGCTGTTATCTAAGCTTACTCTGTGCAAAAATACAGTATTCCTA 974
Db 971 AAAACATTAAAGGAGAGGTTTCAAAAAGCCAAACCTATTCACTACAGGATGGCCCA 1030
Qy 975 TCATACCTGACGAGACAGATTTGATGCTGTGCTCAACAGGCTCGGAAGCTGCGG 1034
Db 1031 TTGTTGCAAGAGATATGATCTTATAGAGATGCGCCAGACTGGAACGAGAAAGCATTTG 1090
Qy 1035 CTTTCTCTCAACAAATTTGCTCATATGATGATGATGAAATTAATCCAGTCTGTTTA 1094
Db 1091 GTTATTTATCTCTGATTTATATCATCTGTGCTCTCAAC-----CCAGCTTAAAG 1141
Qy 1095 AAGAGTGCAGAAACAGAGATTTATTTAGTACCAACTCGAATTTGTCACACAGA 1154
Db 1142 GTCAAGGAATAGACCCGCGCATGTTAGTTCTAATCCACCTCGGAAATGACCTTCAAG 1201
Qy 1155 TTATTTGGAAGCCAGAAATTTTCTTTGGGACTTGTGAAGCTGTGTATATATG 1214
Db 1202 TAGAAGAGATATGTCAAATATTCATAT---AAGGCTTCGAGAGTGTGTATATG 1258
Qy 1215 GGGGAACCCAGCTGGACATTCATTTGACAAATAGTACAGGCTGTATATATATG 1274
Db 1259 GTGTGGAATATAGATGAAACAAATAGAAAGCTTAAAGAGGTATATATATATG 1318
Qy 1275 CTACTCTGGAAGCTGATGATATCATAGCAAGAAAGATTTGGTCTCAACAGATCA 1334
Db 1319 CAATCTCCGGAAGATTAATGATCTGCAATGATTAATCTTCAATCTGAAGATTAAT 1378
Qy 1335 AATACCTAGTTTGTGATGAAGCTGATGCAATGTTGATATGAGTTTGTGTCAGAAATGA 1394
Db 1379 CTTACTTGTGTTTATGATGAAGCAGACAAAGATTTGACATGGGATTTGAACCCAGATTA 1438
Qy 1395 AGAAGTAAATTTCTGCGCCAGGAATGCCATCAAGAAACAGCGCCAAACCTTATGTTCA 1454
Db 1439 TGAAGATTTGT-----TAGATGTGGCCCAAGTATGAGCAAGATTAATATG 1486
Qy 1455 GTGCACTTTTTCAGAGGAATTTCAAGGTTGCTGACAGATTTT---AAAGTCAAAAT 1511
Db 1487 GTGCTATAGCGCTCATTCAGTTATCATGCTTCGACAACTTATTTGAAGAACCAATGA 1546
Qy 1512 ATCTGTTTGTGCTGTTGCAAGTGGGTGAGACATGTAGAAATGTTTACGACAGCCGTT 1571
Db 1547 TTGCTATGTTGATCACTTGGATCTAGTGTCTTAAAGTTCAGGAAGCAAAATATATG 1606
Qy 1572 TCCAGTTGCGCAGTTCTCAAAAAGAAAGAGCTGTTGAAATTTGCAAAACATAGGGG 1631
Db 1607 TAAACACCGAGAGAAAGAAATGAGTCAATGCAAACTTTCTACAGATATGTCAATCA 1666
Qy 1632 ATGAAGAACTATGCTGTTGTTGTAACATPAAGAAAAAGCAGATTTTACTGCAACTTTTC 1691
Db 1667 CAGACAAAGTCAATGCTGTTGTTCTGAAAGAGCTGTTCCGATCTTATCAAGTAC 1726
Qy 1692 TTGTCAAGAAAAATATCACTACAGATATCATGATGAGGGAACAGAGAGCGGG 1751

Db 1727 TAATCTTGGAATATATCTAGTAGCTCTGCATGAGATAGAGAACAGAGATCGGG 1786
QY 1752 AGCAAGCTCTTGAGATTTTCGCTTTGGAAAGTCCAGATCTTGTGCTACTTCAGTAG 1811
Db 1787 AGAAGCATTTAGAGAACTTTAAACAGGCAAGTGAGAAATCAATTGCACTGATCTAG 1846
QY 1812 CTGCGAGAGGGCTGGAATTTGAAAATGTCACATGTTATCAATTTTGATCTCTCTTA 1871
Db 1847 CCTGTAGAGACTGATGTCATGACGTACACATGCTATTAATTTGACTTTCCACGGA 1906
QY 1872 CCATTGATGATATGTCATCGAATTTGGCGCTACTGCTGCTTGGGAATACTGCGAG 1931
Db 1907 ATATTGAAGATAGTACACCGAATAGGGCGACCGGAAGAGAGGAGAGACTGATTT 1966
QY 1932 CAATTCCTTTT 1944
Db 1967 CCATTACACTTT 1979

RESULT 9

US-09-567-995-42
; Sequence 42, Application US/09567995
; Patent No. 6303756
; GENERAL INFORMATION:
; APPLICANT: Martelange, Val,rie
; APPLICANT: De Smet, Charles
; APPLICANT: Boon-Falieur, Thierry
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR
; FILE REFERENCE: L0461/7054
; CURRENT APPLICATION NUMBER: US/09/567,995
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/183,706
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 42
; LENGTH: 2365
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (208)...(2151)
US-09-567-995-42

Query Match 7.1%; Score 158.2; DB 4; Length 2365;
Best Local Similarity 50.1%; Pred. No. 9,2e-36;
Matches 518; Conservative 0; Mismatches 488; Indels 27; Gaps 4;

QY 915 ACAACATTGCTAAGCTGTTAATACTAAGTTACTCTGTCAGAAAATACAGTATTCCTA 974
Db 971 AAAACATTTAAAGAGCGGTTTTCAAAAGCCACACCTATTTCAGTCACAGCATGCGCA 1030
QY 975 TCATCTTGAGAGAGAGATTTGATGCTGCTCAACAGGGCTGAGAGACTGCGG 1034
Db 1031 TTGCTTGCAAGAGATAGATCTTATAGAGTAGCCAGCATGGAACAGGAACAATTTG 1090
QY 1035 CTTTCTCCCAACCAATTTTGCTCATATGATGATGATGAGAAATACATGCGCAATG 1094
Db 1091 GTTATTTAATGCTGATTTATTCATCTGCTCTTCAAC-----CCAGCTTTAAAG 1141
QY 1095 AAGAGTTCAGAGAACAGAGATTTATTTAGACACCACTGAGAAATGTTCAACGAG 1154
Db 1142 GTCAAGAGATAGACCCGCGATGTTAGTTCTAATCCCATTCGGGAATTAGCATTCAG 1201
QY 1155 TTTATTTGAAAGCCAGAAATTTCTTTGAGACTGTGTAAGAGCTGTTGTAATATG 1214
Db 1202 TAGAAGAGATGTTGCAAAATATTCATAT--AAAGGCTTCGAGAGTGTGATATAG 1258
QY 1215 GGGAGACCCAGCTGAGCATTCATTCGACAAATAGTACAAAGCTGTAATATTTATG 1274
Db 1259 GTGCTGAATATGAGATGACAAATAGAGAGCTTTAAAAAGGTGTGATATCATATATG 1318
QY 1275 CTACTCTGAGAGACTGATGATATCATAGCCAAAGAAAGATTGCTCTCAACAGATCA 1334

Db 1319 CAATCCCGGAGATTTGATGATCTGCAAAATGAGTAATTTGCTCAATCTGAAGATATA 1378
QY 1335 AATACTTAGTTTGGATAGAGCTGATCCGATGTTGATATGAGGTTTGGTCCAGAAATGA 1394
Db 1379 CCTACTGTTTATGATGAGACAGCAAGATGTTGACATGGGATTTGAAACCCAGATTA 1438
QY 1395 AGAAGTTAATTTCTTCCCGAGAAATGCCATCAAGAGAACAGGCCAACCTTATGTTCA 1454
Db 1439 TGAAGATTTGT-----TAGATGCGCCCGAGATAGGACAGATTAAGCA 1486
QY 1455 GTGCAACTTTTCCAGAGAGAAATCCAAAGTTGCTGTCAGAGTTT---AAAGTCAATT 1511
Db 1487 GTGCTAATATGCTCATTCAGTTATGCTGCTCGCAATCTTATTTGAAGAACCAATGA 1546
QY 1512 ATCTGTTTGTGCTGTGAGCAAGTGGTGAAGCATGTAGAGATTTAGACAGACCGTTC 1571
Db 1547 TTGCTATGTTGATCATGATGATCTAGTGTGTAAGTTCAGTGAAGCAAAATATTAATTG 1606
QY 1572 TCCAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGCGAAACATAGGGG 1631
Db 1607 TACACCGAGAGAGAGAGAAATGAGTCAATGCAAACTTTCTACAGATATGTCATCA 1666
QY 1632 ATGAAGAACTATGCTCTTTGTTGAACCTAAGAAAAAGCAATTTTACTGCAACTTTC 1691
Db 1667 CAGACAAAGTCAATGCTCTTCTCGAAAAAGCTGTGCGGATCACTTATCAAGTAC 1726
QY 1692 TTTGTCAAGAAAAATATCAACTCAAGTATCCATGCTGATCGGGAACAGAGAGCGG 1751
Db 1727 TAATCTTGGAATATATCAGTAGAGTCTTGCTGATGAGATAGAGAACAGAGATCGG 1786
QY 1753 AGCAAGCTCTTGAGATTTTCGCTTTGGAAAGTCCAGATCTTGTGCTACTTCAGTAG 1811
Db 1787 AGAAGCATTTAGAGAACTTTAAACAGGCAAGTGAATATCTAATTCACATGATCTAG 1846
QY 1812 CTGCGAGAGGGCTGATATGAAAATGTGCAACATGTTATCAATTTTATCTCTCTCTA 1871
Db 1847 CCTGTAGAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1906
QY 1872 CCATTGATGATATGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1931
Db 1907 AATTTGAAGATATCTACACCGAATAGGGCGCAGGAGAGAGAGAGAGTGTGTT 1966
QY 1932 CAATTCCTTTT 1944
Db 1967 CCATTACACTTT 1979

RESULT 10

US-09-328-352-509
; Sequence 509, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 509
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-509

Query Match 5.8%; Score 129.4; DB 4; Length 1191;
Best Local Similarity 47.6%; Pred. No. 1.6e-27;
Matches 521; Conservative 0; Mismatches 556; Indels 18; Gaps 4;

QY 880 TTTTGAAGAGCTATCTCTGTGACAGACTGATATACAACTTGTATAAGCTGTATATC 939
Db 51 TTTTGAACCTTGAATTTTACATCCGCAACTTAAAAAGCGATTGATGCTTTAGGGTTTAC 110

QY 940 TAACTTACTCTGTGCAAAAATACAGTATTCCTATCANAATCTGACAGAGAGATTGAT 999
 Db 111 TCAATGACCCCAATTCAGCAAAAAGTTTAAATATACATTAAGCCGGGACATGATCAAT 170
 QY 1000 GGCTTGTGCTCAAAAGGGGTCTGGGAAGACTGGGGCTTTTCTTCTCAAAATTTGGCTCA 1059
 Db 171 TGGGCGGACAGACAGAAACAGGTAAAGACTGCTGCTTCTTGATTAAGTAAATTAATA 230
 QY 1060 TATGATGATGATGATTAATCACTCCAGTCTTTTAAAGAGTTGACGAAACAGAGTAT 1119
 Db 231 TTGGCTCAATATCCGGTTCAGAGACAGCTTCCGTGGT-----GAGCCTGTGCTTT 284
 QY 1120 TATTGTAGCACTCTGAGAAATTTGTGCAACAGATTTTATTGGAAGCCAGAAATTTTC 1179
 Db 285 AATCTTAGACCTACTGTGAGTTGGATTTGAGATGAAAGTAGCAAAATCTCTCAC 344
 QY 1180 TTTTGGACTTGTGAAGACTGTGTTATATATGAGGGAACCCAGCTGGACATTCAT 1239
 Db 345 AAATTTTCTAATTTACACTAGTACCGCTATTAAGTGGTGTGATTGATGATAGAGAA 404
 QY 1240 TCGACAAATAGTACAAAGCTGT---AATATATTATGTCTACTCTCGAAGACTGATGA 1296
 Db 405 GAAACCACTGATGCTAATTTTGTGATATATGATGTGCAACACAGGGCGTTTAAATGA 464
 QY 1297 TATCATAGCGCAAGAAAAGATTGCTCAAAACAGATCAATACTTATGTTTGGATGAAC 1356
 Db 465 TTTTGTGAAACAAAAGAAAGTTTGGCTCCATCAAAATTTGAAATTTTATGATTAAGTAC 524
 QY 1357 TGAATGATGTGATATGGGTTTGGTCCAGAAATGAAGAAATTAATTTCTTGGCCAG 1416
 Db 525 TGACCGTTTATGGATATGGGCTTTTATCTTGGTAAACGATATGTGCTGTTATCA-- 582
 QY 1417 AATGCCATCAAGGAACAGGCCAACCCCTATGTTCAAGTCACTTTCCAGAGGAAT 1476
 Db 583 ----CCGCTAGAGAAACAGTCAAACTTAATGTTCTGTGCAACATTTAGTATGATGT 638
 QY 1477 TCAAGGTTGGCTGCAAGATTTTAAAGTCAATATCTGTTGTTGCTGTGGAACAT 1536
 Db 639 CTGAATCTGGCAAGACAGTGTTA---TTGAACAGATTAAGTTGAATTAAGACTGA 695
 QY 1537 GGGTGAAGCATGTAGAGATTTGAGACAGACCTTCTCCAAAGTTGGCAAGTTCTCAAAA 1596
 Db 696 ACAAAGACCAATATGATGTGCAACAGGTGTTACGTTGGCTTAAACAAAGATTAATA 755
 QY 1597 AGAAAGCTGCTGAATTTCTGCAAAACATAGGGGATGAAGAACTATGCTTTGTTGA 1656
 Db 756 TCGCTTTTACAGATATTTTACGTGAAGCCAAATGATTAAGATATGATCTTTGCCAA 815
 QY 1657 AACTAGAAAAAGCAGATTTTACTGCAACTTTCTTTGTCAAGAAAAATATCAATAC 1716
 Db 816 TCGCCGTGATCAGGTACGTGCTTTATGACATTTGAAAAAAGATGATATTAAGTGG 875
 QY 1717 AAGATCCATGTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTTCGCTT 1776
 Db 876 GATGCTATCTGTGAATTTCTCAAGATTAAGCTTTAAAAATGTTAGACAGATTTTAGCA 935
 QY 1777 TGAAGATGCCAGTTCTTGTGCTAATTCAAGTCTGCAAGGGCTGGATATGAAAA 1836
 Db 936 AGGCAAAACATACATGATGATGCAACGAGATTTGCTGCTGATTAATCATGTAGACGG 995
 QY 1837 TGGCAACATGATTAATTTGATGATCTCTTACCAATTGATGATGATGATGATGATGAT 1896
 Db 996 TGTATCATGATGATGATTAATTTTACATTTGCTGAGAGCTGTGATGATTAATGCTGAT 1055
 QY 1897 TGGGCGTACTGTGCTGTGGAATACTGGCAGAGCAATTTCTTTTGTGATCTGAATC 1956
 Db 1056 TGGTTCGACAGCTGCTGACAGGGCACAAGGTGTAAGTATTAAGTTTCTATCGAAGATGA 1115
 QY 1957 GGTAAACATTTAGC 1971
 Db 1116 TGCTTTCTATTAC 1130

RESULT 11
 US-09-976-594-213
 ; Sequence 213, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 213
 ; LENGTH: 3760
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURES:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 2173757CB1
 US-09-976-594-213
 Query Match 5.8%; Score 129; DB 4; Length 3760;
 Best Local Similarity 48.4%; Pred. No. 3,9e-27;
 Matches 539; Conservative 0; Mismatches 535; Indels 39; Gaps 5;
 QY 844 AGTGTCTGACATGATGACACACACAGCAATTTGACTTTTGAAGAACTATCTGTCTCA 903
 Db 1236 AGTTAAAGAAAGAGTTGCCCAACCAATTAATCTGGGTCAAGTGGAAATTTCCAT 1295
 QY 904 GACACTGAATATACAACTTGTCTAAAGCTGTTATCTACTACTCTCTGCAAAAATA 963
 Db 1286 GAATATCTTAATTTCTCTCAAGAAAGATGCTATGAAAGCCAGCCCATCCAAACCA 1355
 QY 964 CAGTATCTTATCATCTTCAAGAGAGCAAGTTGATGCTTGTCTCAAAAGGCTGTG 1023
 Db 1356 AGTATTCCTGCTAATATGCTGTGACAGAGATTTGATTTGATTTGCCAAACAGAAAGTGG 1415
 QY 1024 GAATCTGCGGCTTTTCTCTACCAATTTTGGCTCATATGATGATGATGATGATGATGAT 1083
 Db 1416 AAGACCAATGCTTTTCTGTGCTGCTGATTTAGACATATGATGATGATGATGATGATGAT 1463
 QY 1084 CAGTCTGTTTAAAGATTGACAGAAACAGAGTATTTATTTGACCAACCTGAGAAAT 1143
 Db 1464 GAGGTCAATTAGAGAGAGAGAGGGGCCAATAGCTGTCATCATGATCTCAACTGAGACT 1523
 QY 1144 GGTCAACCAATTTATTTGGAAGCCAGAAATTTCTTTTGGACCTTGTATAGACTGT 1203
 Db 1524 GGCCTTACAGATTAATTAAGAGTGTAAAGATTTTCCAGACTTTGGACTTGAAGTGT 1583
 QY 1204 TGTATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAGAGCTGTAA 1263
 Db 1584 CTGTGTTTACGAGAGAAACAGAAATCAATGACAGATTTGCTGACTGAAAGAGTGTCTGA 1643
 QY 1264 TATATATGTGCTACTCTCTGGAAGACTGATGATATCATAGCAAGAAAGAAATGTGCT 1323
 Db 1644 AATATATGTTTGCACACCTGCTGCAATGATGACATTTGACCGCTAACAGTGTGGGT 1703
 QY 1324 CAACAGATC-----AATACTAGTTTGGATGGAAGCTGATGCTGTTGATAT 1374
 Db 1704 CACAAATCTTGAAGATGACATATGTTTATTTAGTGAAGACAGACGAATTTTGAAT 1763
 QY 1375 GGGTTTGTCCAGAAATGAAGATTAATTTCTTCCAGGAATGCCATCAAGAGACA 1434
 Db 1764 GGGTTTGAACCCAGGTCAATGCGCATGCGATTAATGTTGCT-----CCTGA 1811
 QY 1435 GCGCAACCTTATGCTCACTGATGCAATTTTCCAGAGGAATTCAAAGTTGCTGTGAGA 1494
 Db 1812 TCGACAGACGTTATGTTTCACTACTTCCCAAGCTATGAGAGCTTTGCTGTGAG 1871
 QY 1495 GTTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGAGTGGGTGACATGATGAGA 1554

Query Match	5.6%;	Score 125.6;	DB 4;	Length 1830121;
Best Local Similarity	47.3%;	Pred. No. 1.3e-24;		
Matches 506;	Conservative	0;	Mismatches 534;	Indels 30;
				Gaps 3;

QY	881	TTTGAAGAAGCTATCTCTGTGACAGACATGAAATAACAATTGCTAAAGCTGGTAAACT	940
Dp	444736	TTTGAACANTGATCTTCTCTCTAGCTTTAAAGGACCTGGAGAAAAAGTTATCC	444677
QY	941	AAAGCTTACTCTGTGCAAAAAATACGATTTCTATCATCTTCAGACGAGATTTGATG	1000
Dp	444676	CGCCCAACGCATTTCTAAATGGAGGCCATTTCTGCGGAAATGGAAGAGATGATTTA	444611
QY	1001	GCTTGTGCTCAACACAGGCTGTGAGAAAGATGCGGCTTTTCTCTCACAATTTTGGCTCAT	1060
Dp	444616	GGCTGGCAACACCGGGAACAGGAAACGTGCTTTTATTTACTCTGGCCTACACAT	444555
QY	1061	ATGATGATGATGGAATTAATCTGCAGTCTTTTAAAGATTGACGAAACCGAGTGATTT	1120
Dp	444556	TTTATGATTAATCAACGCGGTAAACGAGGCC-----ACGACGATTTTGG	444512
QY	1121	ATTGTGACCAACATCGAAGATTGGTCACACGATTTTATTTGGAGCAGAAAAATTTCT	1180
Dp	444511	GTTATTAAACCAACCGGTGAATCTGGCAATGCAAGTGTGTAACAACGGAAGATTAGCG	444455
QY	1181	TTTGGACCTTGTGTAAAGACTGTGTTATATATGAGGGAACCCAGCTGGGACATTCATT	1240
Dp	444451	CAGTTCACCCATTTAAATATATTCGCAATTAACGCTGGCGTGGCCTATCAAAATCACGGT	444392
QY	1241	CGACAAATATGTCAAGAGCTGTATATATATATGTGTACTCCGGAAGCTAGATGATTC	1300
Dp	444391	GATGTATTCAATACCAATCAATCAATTTGTGTGGGCTAGCCAGGCCGTTTGTTCATATC	444332
QY	1301	ATAGGCAAGAAAAGATTGCTCTCAACACAGATCAATACTTAGTTTGGATGAAGCTGAT	1360
Dp	444331	ATTAAAGAAAGAAATTTTATTTGATTCGCTTCGTTCCGTTGAAAGCTGATTTTGTGAAAGCCGAT	444272
QY	1361	CGCATTTTGGATATGGGTTTGTGTGTCAGAAATGAAGAAGTAAATTTCTTGCCAGGAATG	1420
Dp	444271	AGAAATTTGCAATATGGGATTTTGGGCAAGATGGGAAAAAATTCGACGTGAAACCCGTTGG	444212
QY	1421	CCATCAAGGAACAGCGCCAAACCTTATGTTCACTGCAACCTTTTCCAGAGAAATTCAA	1480

Db 444211 CGAAAA-----CAAACTTGTGTTTCTGTGCAACTTAGAAGAGAGATTATG 444164

Qy 1481 AGGTGGCTGCAGAGATTTTAAAGCAATATACGCTTGTGTGTGACAAGTGGT 1540

Db 444163 GTGCAATTCGGGAGCGTTTATTGAATATCTGTGAAGTAGATGGGAAACCAAGTGC 444104

Qy 1541 GGAGCATGTAGAGATGTTCAAG--GACCGTTTCCAGTTGGCCAGTTCTCAAAAAGA 1597

Db 444103 CGGAAAGAAAAAATCAATCAATGATATTATACGAGACAGACATGAAACAAATC 444044

Qy 1598 GAAAAGCTGTGAATAATCTGTGGAACAATAGGGATGAAAGAACTATAGCTTTGTGGA 1557

Db 444043 AAATTGCTGGCGGCTTTATTGAACGTAGAGAAAGTAACCGTGGAAATGTGTTATTGCT 443984

Qy 1658 ACTAGAAAAAAGCAGATTTTACTGCAACTTTCTTGTGCAAAAAAATATCAACTCA 1717

Db 443983 CGTCGTGAAGATGACGGAACGTTCCTCCAAACATTTGGTAAACGAGGCATTCGTGCGCA 443924

Qy 1718 AGATTCATGATGATCGGGAACAGAGAGCGGAGACAAGCTCTTGAGATTTTCGCTT 1777

Db 443923 TATTTAGAAGGGAATGCAACAACCTCAAGTAAACATGCAATGTATTAATTGAATCA 443864

Qy 1778 GGAAGTGGCCAGTTCTTGTGTGCTACTTCACTAGCTGCCAGAGGCGCTGCAATTTGAAAT 1837

Db 443863 GGAATTGTGACGTAATTTGTTGCAACAGAGTGGCTCAGCTGATATGATTAATGACGAT 443804

Qy 1838 GTGCAACAATGATATCAATTTGATCTTCTCTTCAACATGATGATATATGTTCAACAATT 1897

Db 443803 GTTAGCAAGTATGAAATTTTGAATTGCTTATGTCCTATATGCGGATCTTAATTGCAATGAAAT 443744

Qy 1898 GGGCGTACTGTCGTGTGGAGAAATACGCGGAGACCAATTTCTTTTGA 1947

Db 443743 GGAACGTACCGCGAGCTGGCAAAAAAGCACGGCAGCTCTTTTGTGCA 443694

RESULT 14
 US-09-643-990A-1/c
 Sequence 1, Application US/09643990A
 Patent No. 6528289
 GENERAL INFORMATION:
 APPLICANT: Robert D. Fleischmann
 Mark D. Adams
 Owen White
 Hamilton O. Smith
 J. Craig Venter
 TITLE OF INVENTION: The Nucleotide sequence of
 the Haemophilus influenzae Rd Genome, Fragments
 Thereof, and Uses Thereof
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville,
 STATE: MD
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS V6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/643,990A
 FILING DATE: 23-Aug-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787
 FILING DATE: 1995-04-21
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover

```

REGISTRATION NUMBER: 40.302
REFERENCE/DOCKET NUMBER: P8186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      5.6%  Score 125.6;  DB 4;  Length 1830121;
Best Local Similarity 47.3%;  Pred. No. 1.3e-24;
Matches 506;  Conservative 0;  Mismatches 534;  Indels 30;  Gaps 3;

QY      881  TTGAAGAGCTAATCTCTGTCAAGACATGATATACAACTTGTCTAAAGCTGTTACT 940
Db      444736  TTGAACAATTCATCTTTCTCCGAGCTTTTAAAGCACTTGAAAAAAGTTATTC 444677

QY      941  AAGTTTCTCTGCGAAAAATACAGTATCTCTATCATCTTCGAGAGCAGATTGANG 1000
Db      444676  CGCCCAACAGCTATTCMAATGGAAGCCATTCTCGCATGGAAGAGATGATATTA 444617

QY      1001  GCTTGTCTCAACAGAGGTGTGGAGAGCTGCGCTTTTCTCTACCAATTTGGCTCAT 1060
Db      444616  GGCTCGGACCAACAGGAAACAGGAAACGTCGCTTTTATTAACCTGGCTACAACT 444557

QY      1061  ATGATGATATGAGATTAATCTCCAGTCGTTTAAAGAGTGCAGAACCAAGTGTTAT 1120
Db      444556  TTATGTATATATCCAGCCCTTAAACAGAGCC-----ACACCTATTTTG 444512

QY      1121  ATTGTACCACTCACTGAGAAATGTGTCAACAGATTATTTGGAGCGAAATTTTCT 1180
Db      444511  GTATTAACACCAACCCGTGACTGCAATGCAATGCTCAACAGCGAAGATTATGCG 444452

QY      1181  TTTGGACTTGTGTAGAGCTGTTGTATATATGTGGGGAACCCAGCTGGGCAATTCAT 1240
Db      444451  CAGTTCAACCATTAATATATTTGGACATATACAGTGTGCGTATCAAAATCAGGT 444392

QY      1241  CGACAAATAGTACAGGCTGTATATATATGTGTCTCTCTGAAAGCTAGATATC 1300
Db      444391  GATGTATTCAAATCAATCAATCAAGATTGTGTGTGCTACGCAAGCCGTTTGTGCAAT 444332

QY      1301  ATAGCAAGAAAGATTTGCTCAACAGATCAAAATCTAGTTTGTGATGAGCTGAT 1360
Db      444331  ATTAAAGAAAGAAATTTGATGTCGCTTGCCTTGAATGCTGATTTTGAATGAGCGAT 444272

QY      1361  CGCATTTGATATGAGTTTGTGCTCAAGAAATGAAGATTAATTTCTTCCAGGAATG 1420
Db      444271  AGAATGTTGAATTTGGAATTTGGCAAGATGCGAAAAAATTCAGCTGAACCCGTGTG 444212

QY      1421  CCATCAAGAAAGACAGCCCAACCCCTATGTTCAAGTCACTTTCCAGAGAAATTCAA 1480
Db      444211  CGAAAA-----CAAATTTGTTTCTTTTCTGCAACCTTAGAAGAGATTAATG 444164

QY      1481  AGGTTGGCTGCAAGTTTAAAGTCAATTAATCTGTTGTGCTGTGTGCAAGTGGGT 1540
Db      444163  GTGATTTCCGCGGCGCTTATTTGAATGATCCGTGAAGATGATGCGGAACCAAGTCCG 444104

QY      1541  GAGCATGTAGAGATGTTGACA---GACGCTCTCCAAATTTGGCCAGTTCTCAAAAAG 1597
Db      444103  CGTGAAGAAAAAATCAATCAATGATATTAACGCAACGCAATGAACCAAAATC 444044

QY      1598  GAAAAGCTGTTGAATTTGCGAAACATAGGGAGTAAAGAACTATGTCTTTTGTGAA 1657
Db      444043  AAATTTGCTCGCGCTTTATTTGAATGAGAAAGTAAACCGTGAATTTGTTATTCGT 443984

QY      1658  ACTAAGAAAAAGAGATTTTACGCAACTTTCTTTTGTGTCAAAAAAATATCACTAAC 1717
Db      443983  CGTGTGAAGATGACGTGAACCTTTCCGAACATTTGGGTAAACGAGGCAATTCGTTCCG 443924

```

```

QY      1718  AGTATCCATGTGTATCGGGAACAGAGAGCGGGAAGCAAGCTTTGGAGATTTGCTTT 1777
Db      443923  TATTTAAGAGGGGAATGSCAACAACTCAAGCTAACATGTCATGATTAATTAATCA 443864

QY      1778  GGAAGTCCCGAGTCTTGTGTCTACTTCACTAGTGTCCAGAGGCGTGTATTAATAAT 1837
Db      443863  GGTATTTGAGCGGATTTGTTGCAACAGATGTGCTGACAGTGTGATTAATGACAT 443804

QY      1838  GTGCAACATGTATATATTTGATTTCTTCTTACCATGATGAATATATGTCATGATTT 1897
Db      443803  GTAGCCACGAGTAAATTTGATTTGCCATATGTGGATTAATTTGCAATGAAAT 443744

QY      1898  GGGGCTACTGTCGTTGTGGGAATATCTGCAAGCAATTTCTTTTGA 1947
Db      443743  GGACGTACCGCGGAGCTGGCAAAAAAGCAAGGCAAGTCTCTTTGTGCA 443694

RESULT 15
PCT-US96-05320A-894
; Sequence 894, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 122, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 894:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US96-05320A-894

```

Query Match 5.6%; Score 123.6; DB:5; Length 1254;
Best Local Similarity 47.1%; Pred. No. 7.9e-26;
Matches 495; Conservative 0; Mismatches 539; Indels 18; Gaps 3;

QY 921 TTCTAAGCTGTTTACTAAGCTTACTCTGTGTGCAAAAACAGTATTCCTATCAATAC 980
DB 80 TGGCAAAAAGGCTTTGATTTTGTACCCCAATTCAGCTTATCTCTGCTATCAGTT 139
QY 981 TTGCAGAGCAGATTGATGAGCTTGTCTCAACAGGCTGTGGAAGACTGCGGCTTTTC 1040
DB 140 TAAATGACGAGATGTGCGACGACAGCTCAAACTGTACAGCAAGCAATGCGCTTTT 199
QY 1041 TCCCTACCAATTTGGCTCATGTGATGATGAAATTAACGCCAGTCGTTTAAAGAT 1100
DB 200 TAACGGCTACTTTTCACCACTTTTAACACCAAGA-----TCCCTAATCTTAAAT 250
QY 1101 TGCAGAACCCAGCTGATATTAATGTAGCAACCACTGAGATTTGTCAACAGATTAT 1160
DB 251 ATCTCAACCCAGAGCTTTGATTTTGAACCTTACCTGAGAAATTAAGCGTACAGATTAG 310
QY 1161 TGGAGCCAGAAAATTTCTTTTGGAGCTGTGTAAAGCTGTGTATATATGAGGAGAA 1220
DB 311 ATGACGAGATTTCTGCAAAAAGCGATGATTAAGACCGCACTTGCTATGAGTGGCG 370
QY 1221 CCGAGCTGGACATTGATTCGACAAATGATACAGGCTGTATATATATGTGCTACTC 1280
DB 371 ATGTTATGATTAACACTACAGCGATGAGCTGCGTCATATTGATTGATGACGA 430
QY 1281 CTGGAAGACTGATGATATCATATGAGCAAAAAGATTTGTCTCAACAGATCAAAATCT 1340
DB 431 CCGGGCGAGTCACTGATTTATGTGAACAAAGGCGTAAATGTTAGATGAATCCAAAGTTG 490
QY 1341 TAGTTTGGATGAGCTGATCGATGTGTGATATGAGTTTGTGTCAGAAATGAAGAGT 1400
DB 491 TCGTGTAGATGAAGCAGATCGAATGTTGATCTTGGGTTTATCCGTGATATCTGTATT 550
QY 1401 TAAATTTCTGCCAGGATGCGCATCAAGAAACAGCGCAACCTTATGTCAATGCA 1460
DB 551 TATTTGGTAA-----ATGCCCGCTCCGCAAGCTGTTAAAGATGTTATTTTCAAGCA 604
QY 1461 CTTTTCAGAGGAATTCAAAGSTTGGCTGACAGATTTTAAAGTCAAAATTAATCTGTTG 1520
DB 605 CGCTTTCTTAATAAGCGCGTAATTAGCAATTGAAGATGATGAATGAACCTGA--ATA 661
QY 1521 TTGCTGTGAGCAAGTGGGTGGAGCANTAGAGATGTGACAGACCGTTCTCCAAAGTTG 1580
DB 662 TTGAAAATGAACAGAACCAAAAACAGGGCACCGAATTAAGAGACATTTTATTCAT 721
QY 1581 GCCAGTTCATAAAGAGAAAAGCTGTTGAATTCGCAAAACATAGGGGATGAAGAA 1640
DB 722 CTAAATCAGGATAAATGGCACTTCTTAACCTTAATGAAAGTGAATGGCTGAACGCT 781
QY 1641 CTATGCTCTTTGTTGAAACTAAGAAAAAGCAGATTTTACTGCACTTTCTTTGTCAAG 1700
DB 782 GTATTTGATTTGGAAATACGAACATGTTGTGAAGAAATTTGGGGCTAATTTGGCGGCTG 841
QY 1701 AAAAATATCAACTACATGATTCATGATCGGAGACAGAGAGCGGAGCAAGCTC 1760
DB 842 ATGGGCAATGCTGCGCTTACTGACGTGGAGATGACACAGAAAAACGTTATCGTTAT 901
QY 1761 TTGGAATTTTCCGCTTTGGAAGTGGCCAGTCTTGTCTTACTCAGTACGCTGCCAAG 1820
DB 902 TAAACCAATTTACTGATGATTTGATTAATTTTGTGCAACAGATGAGTGGCTGCTG 961
QY 1821 GCGTGGATTTGAAAAATGTCACATGTTATCAATTTGATCTTCTTACCATTTGATG 1880
DB 962 GCTTGCATATTTCTGATGTACGCAATGTTTCAATTAATGATTAACCGATGATCGAAG 1021
QY 1881 AATATGTTCAATGAAATGGCGGCTGCTGCTGTGTGGAATACCTGGCAGAGCAATTTCT 1940
DB 1022 ATATATGTTCAACGAATGGGCGTACTGACGAGCAAGGGAAGTGTGTTGATTAAGTT 1081
QY 1941 TTTTGTATCTGATGCAATACCAATTGCA 1972

DB 1082 TCGCTGTGAAGATATGCGATGAAATTTACCA 1113

Search completed: July 27, 2004, 10:52:53
Job time : 189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 02:46:37 ; Search time 880 Seconds
(without alignments)
10736.355 Million cell updates/sec

Title: US-09-714-865B-1

Sequence: 1 acttgaaagtcaccatggggg.....aagtcctgtggttttgatgca 2224

Scoring table: IDENTITY_NUC.

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
1: geneseqn29jan04.*
2: geneseqn1980s.*
3: geneseqn1990s.*
4: geneseqn2000s.*
5: geneseqn2001as.*
6: geneseqn2001bs.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length		DB ID	Description
			DB	Length		
1	2224	100.0	2224	5	AAD06354	Aad06354 Human vaa
2	1601	72.0	1984	9	ADB63498	Adb63498 Human cdn
3	871.4	39.2	1006	3	AAA44726	Aaa44726 Human sec
4	442.4	19.9	677	4	AAS40979	Aas40979 cDNA enc
5	435.2	19.6	667	7	ABZ20094	Abz20094 Group II
6	386.4	17.4	2451	7	ABX63480	Abx63480 Human CDN
7	384.8	17.3	4423	7	ABX63951	Abx63951 Human act
8	383.2	17.2	4416	2	AAV69633	AAV69633 Dead Box
9	383.2	17.2	4416	7	ACA64934	ACA64934 Human DB
10	383.2	17.2	4416	9	ADD15196	ADD15196 Human dea
11	371	16.7	2187	6	AB199243	Ab199243 Mouse isc
12	371	16.7	3158	7	ACD13409	ACd13409 Human DNN
13	369.4	16.6	3408	2	AAV69632	AAV69632 Dead Box
14	369.4	16.6	5321	3	AAA29207	AAA29207 Human DBX
15	369.4	16.6	5322	2	AAV69631	AAV69631 Dead Box
16	369.4	16.6	5322	7	ABZ72018	ABZ72018 Human dea
17	356	16.0	591	4	AAS41534	Aas41534 cDNA enc
18	356	16.0	591	4	AA101169	AAa101169 Human rep
19	356	16.0	591	4	ABP96630	ABP96630 Human tee
20	327	14.7	2019	6	ABR32277	ABr32277 Candida a
21	313.8	14.2	2236	4	ABE06875	ABl06875 Drosophi
22	291.4	13.1	2320	3	AAC37735	AAc37735 Arabidop
23	283.4	12.7	2244	3	AA338986	AAc38986 Arabidop

[illegible]

ALIGNMENTS

RESULT 1
ADD00004

ID	AAD06354	standard; cDNA; 2224 bp.
XX	AAD06354;	
AC		
XX		
DT	10-AUG-2001	(first entry)
XX		
DE	Human vasa CDNA.	
XX		
KW	Human; vasa; germ cell; gonad development; therapy; cancer; oral; brain;	
KM	ovarian; biliary tract; breast; pancreas; prostate; colorectal; cervical;	
KM	colon; lung; testis; renal; thyroid; oesophageal; endometrial; gastric;	
KW	skin; liver; lymphoma; melanoma; glioblastoma; neuroblastoma; neoplasm;	
KM	metalloblastoma; choriocarcinoma; squamous cell carcinoma; leukaemia;	
KM	acute lymphocytic; myelogenous; multiple myeloma; Paget's disease;	
KM	osteosarcoma; Acquired immune deficiency syndrome; AIDS; Bowen's disease;	
KM	leiomyosarcoma; rhabdomyosarcoma; liposarcoma; Kaposi's sarcoma;	
KW	fibrosarcoma; tumour; testicular; seminoma; ovarian; dysgerminoma;	
KM	teratoma; mediastinal; intracranial; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	14..2188
FT		/tag= a
FT		/product= "Human vasa protein"
FT	misc_feature	14..2185
FT		/tag= b
FT		/note= "this region is specifically claimed as SEQ ID NO
FT		15 in claim 3 of the specification"
XX		
PN	WO200136445-A1.	
XX		
PD	25-MAY-2001.	
PF	16-NOV-2000; 2000WO-US031485.	
XX		
PR	18-NOV-1999; 99US-0166394P.	
XX		
PA	(BGM) BRIGHAM & WOMENS HOSPITAL INC.	
XX		
PI	Castillon DH;	
XX		
DR	WPI; 2001-355606/37.	
DR	P-PStDb; AAE02417.	

XX Novel vasa polynucleotides useful in the diagnosis or treatment of
PT conditions characterized by aberrant expression and/or presence of mutant
PT forms of vasa polynucleotides or polypeptides.

XX Claim 1; Page 52-53; 66pp; English.

CC The present sequence is human vasa cDNA that has germ cell specific
CC expression and is believed to play a determinative role in gonad
CC development. Germ cells are specialised to produce haploid gametes in
CC multicellular organisms. Vasa is useful in the diagnosis or treatment of
CC conditions characterised by its aberrant expression and/or the presence
CC of its mutant forms. The conditions include cancers such as biliary
CC tract, brain, breast, colon, ovarian, pancreas, prostate, colorectal,
CC oral, liver, lung, skin, basocellular, testis, renal, thyroid, cervical,
CC endometrial, oesophageal and gastric, lymphomas, melanomas,
CC glioblastoma, neuroblastomas, medulloblastomas, choriocarcinoma,
CC squamous cell carcinoma, haematological neoplasms, acute lymphocytic and
CC myelogenous leukaemia, multiple myeloma, Acquired immune deficiency
CC syndrome (AIDS) associated leukaemias, intraepithelial neoplasms, Bowen's
CC disease, Paget's disease, sarcomas such as leiomyosarcoma,
CC rhabdomyosarcoma, liposarcoma, fibrosarcoma, Kaposi's sarcoma and
CC osteosarcoma, tumours such as testicular tumour (eg. seminoma), ovarian
CC tumour (eg. dysgerminoma or teratoma) and tumour of an extragonadal
CC tissue (eg. mediastinal or an intracranial tumour)

XX Sequence 2224 BP; 678 A; 385 C; 550 G; 611 T; 0 U; 0 Other;

Query Match 100.0%; Score 2224; DB 5; Length 2224;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTTGAAGTCACATGGGGGATGAGATGGGAAAGCAAAATCAACCTCATATGCTTC 60
DB 1 ACTTGAAGTCACATGGGGGATGAGATGGGAAAGCAAAATCAACCTCATATGCTTC 60
QY 61 CTATGTTCCCATATTTGAGAAGATAGTATTTCTGAGAAATAGAGACATTTTACAG 120
DB 61 CTATGTTCCCATATTTGAGAAGATAGTATTTCTGAGAAATAGAGACATTTTACAG 120
QY 121 GACTCCAGCTTCTCATCAGAAATGATGATGACCTTCTCGAAGATCATTTTCAGAA 180
DB 121 GACTCCAGCTTCTCATCAGAAATGATGATGACCTTCTCGAAGATCATTTTCAGAA 180
QY 181 AAGTGAATTTGCTCTGGGGCGAATTTTGGAAACAGAGATGCTGGATGTATTAAGCG 240
DB 181 AAGTGAATTTGCTCTGGGGCGAATTTTGGAAACAGAGATGCTGGATGTATTAAGCG 240
QY 241 AGATTAATATCATCAATGGGTGTTTGGAGTTGAAAGAGTTTGGAAACAGAGCTT 300
DB 241 AGATTAATATCATCAATGGGTGTTTGGAGTTGAAAGAGTTTGGAAACAGAGCTT 300
QY 301 TTCAAAACAGAGGTTGAGATGATGATGCTGCTGTTCTGAGAGAGCTTAATGA 360
DB 301 TTCAAAACAGAGGTTGAGATGATGATGCTGCTGTTCTGAGAGAGCTTAATGA 360
QY 361 CTGCGAAGATTAATCCACAACGGAACAGAGGTTTTCAGAGAGCGGCTATCGAGATG 420
DB 361 CTGCGAAGATTAATCCACAACGGAACAGAGGTTTTCAGAGAGCGGCTATCGAGATG 420
QY 421 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTATTCGAGGTTG 480
DB 421 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTATTCGAGGTTG 480
QY 481 CCGTGAAGAGTTTGTCTAGAAAGTCCAATATATGACTTAAGCCAGAGGAATGTATCA 540
DB 481 CCGTGAAGAGTTTGTCTAGAAAGTCCAATATATGACTTAAGCCAGAGGAATGTATCA 540
QY 541 GGGCATGTGGTGGCTTTTGGTTCTAGAAACCAAGATTTAAGTGCAACAGTAAATG 600
DB 541 GGGCATGTGGTGGCTTTTGGTTCTAGAAACCAAGATTTAAGTGCAACAGTAAATG 600
QY 601 TACTTCTCAACAGAGAGTGCAGTGAAGTGAACGAGGTGTTCAAAAGTTTAAATGA 660

DB 601 TACTTCTCAACAGAGAGTGCAGTGAAGTGAACGAGGTGTTCAAAAGTTTAAATGA 660
QY 661 AGAATTAATTAACAGGCTTGGAAAGATTTCTTGAAGTCAGAACAGAGAGAGAAAG 720
DB 661 AGAATTAATTAACAGGCTTGGAAAGATTTCTTGAAGTCAGAACAGAGAGAGAAAG 720
QY 721 TAGTATTAATCAAGGACCAAAAGTACATATACCCCTCCACCTGAGATGAGGA 780
DB 721 TAGTATTAATCAAGGACCAAAAGTACATATACCCCTCCACCTGAGATGAGGA 780
QY 781 CTCCATCTTTGACATTTTCAAGAGCATTAATCTTGACAAATACGACTATTTCTGT 840
DB 781 CTCCATCTTTGACATTTTCAAGAGCATTAATCTTGACAAATACGACTATTTCTGT 840
QY 841 GGAATGTCTGACATGATGACACCAAGCATTTCTGACTTTTGAAGAGTATCTCTG 900
DB 841 GGAATGTCTGACATGATGACACCAAGCATTTCTGACTTTTGAAGAGTATCTCTG 900
QY 901 TCAGACTGAATTAACAACATTTGCTAAAGCTGTTATATCAAGTTTACTCTGTGCAAA 960
DB 901 TCAGACTGAATTAACAACATTTGCTAAAGCTGTTATATCAAGTTTACTCTGTGCAAA 960
QY 961 ATACAGATTTCTTCAATACCTTGAGAGACGAGATTTGATGGCTTGTCTGCAACGGCTC 1020
DB 961 ATACAGATTTCTTCAATACCTTGAGAGACGAGATTTGATGGCTTGTCTGCAACGGCTC 1020
QY 1021 TGGGAAGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGAGATTAAC 1080
DB 1021 TGGGAAGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGAGATTAAC 1080
QY 1081 TGCCAGTCTTTTAAAGATTGACAGAACCAAGATGATTTATGTATACCAACTGAGA 1140
DB 1081 TGCCAGTCTTTTAAAGATTGACAGAACCAAGATGATTTATGTATACCAACTGAGA 1140
QY 1141 ATTGCTAACAGATTTATTTTGAAGCCAAATAATTTCTTTGGGACTGTGTAAGAGC 1200
DB 1141 ATTGCTAACAGATTTATTTTGAAGCCAAATAATTTCTTTGGGACTGTGTAAGAGC 1200
QY 1201 TGTGTTATATATGAGGGAACCAAGCTGGACATTCATTTGCAAAATAGTACAGGCTG 1260
DB 1201 TGTGTTATATATGAGGGAACCAAGCTGGACATTCATTTGCAAAATAGTACAGGCTG 1260
QY 1261 TAAATATATATGCTCTCTCTGGAAGACGATGAGATATATGAGCAAAAGAAATTTGG 1320
DB 1261 TAAATATATATGCTCTCTCTGGAAGACGATGAGATATATGAGCAAAAGAAATTTGG 1320
QY 1321 TCTCAACAGATCAAAATATAGTTTGAATGAAGCTGATCGCATGTGTAATGGGTT 1380
DB 1321 TCTCAACAGATCAAAATATAGTTTGAATGAAGCTGATCGCATGTGTAATGGGTT 1380
QY 1381 TGGTCCAGAAATGAAGAGTTAATTTCTTCCCAAGAAATGCCATCAAAAGAACAGGCCA 1440
DB 1381 TGGTCCAGAAATGAAGAGTTAATTTCTTCCCAAGAAATGCCATCAAAAGAACAGGCCA 1440
QY 1441 AACCTTATGATGATGCAACTTTTCCAGAGAAATTCAAAGGTTGGCTGCAAGTTTTT 1500
DB 1441 AACCTTATGATGATGCAACTTTTCCAGAGAAATTCAAAGGTTGGCTGCAAGTTTTT 1500
QY 1501 AAAGTCAAAATTAATCTGTTGTTCTGTGGAACAAGGGGTGAGCATGTAGAGATGTCA 1560
DB 1501 AAAGTCAAAATTAATCTGTTGTTCTGTGGAACAAGGGGTGAGCATGTAGAGATGTCA 1560
QY 1561 GCAGACCGTTCTCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGGG 1620
DB 1561 GCAGACCGTTCTCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGGG 1620
QY 1621 AAACATATGAGGAGTAAAGAACTATGATCTTTGTTGAACCTTAAGAAAAGACAGATTTAC 1680
DB 1621 AAACATATGAGGAGTAAAGAACTATGATCTTTGTTGAACCTTAAGAAAAGACAGATTTAC 1680
QY 1681 TGCAACTTTTCTTGTGCAAGAAAATATCAACTTAACAGATTCAGTGTGATGCGGAGAC 1740


```
Db 1681 TGCAATCTTTCTTGTGCAAGAAAAATATCAACTACAGATCATGATCGGAGACA 1740
Qy 1741 GAGAGAGCCGGGAGCAAGCTTTTGAGATTTTGGCTTTGGAAGAGCCAGTTCTGTGTC 1800
Db 1741 GAGAGAGCCGGGAGCAAGCTTTTGAGATTTTGGCTTTGGAAGAGCCAGTTCTGTGTC 1800
Qy 1801 TACTTCAGTAGCTGCCAGAGGGCTGATATTTGAAAATGTGCAACATGTTATCAATTTTGA 1860
Db 1801 TACTTCAGTAGCTGCCAGAGGGCTGATATTTGAAAATGTGCAACATGTTATCAATTTTGA 1860
Qy 1861 TCTTCTCTTCAACATGATGATATTTGTTATTCGATTTGGGCGTACTGCTGTGTGGGA 1920
Db 1861 TCTTCTCTTCAACATGATGATATTTGTTATTCGATTTGGGCGTACTGCTGTGTGGGA 1920
Qy 1921 TACTGAGAGCAATTTCTTTTGTGATTTGATCGATTAACCATTTAGCAGGCTCT 1980
Db 1921 TACTGAGAGCAATTTCTTTTGTGATTTGATCGATTAACCATTTAGCAGGCTCT 1980
Qy 1981 AGTAAAGATATGACAGATGCTCAACAGATGTTCTGCAATGTTGGAAGAAATTCCTT 2040
Db 1981 AGTAAAGATATGACAGATGCTCAACAGATGTTCTGCAATGTTGGAAGAAATTCCTT 2040
Qy 2041 TAGTACATACATTCCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGAT 2100
Db 2041 TAGTACATACATTCCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGAT 2100
Qy 2101 TACAGAGAGGAGCAAGCACTTTGAAACACAGCTGCTGTTCTTCTTACAGAGCTCCAA 2160
Db 2101 TACAGAGAGGAGCAAGCACTTTGAAACACAGCTGCTGTTCTTCTTACAGAGCTCCAA 2160
Qy 2161 TCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Db 2161 TCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Qy 2221 TGCA 2224
Db 2221 TGCA 2224
```

```
RESULT 2
ADB63498
ID ADB63498 standard; cDNA; 1984 BP.
XX
AC ADB63498;
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone TEST120227380.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 20..11747
FT /tag= a
FT /product= "Clone TEST120227380 protein"
```

```
EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002BP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
```

```
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isoo Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
DR MPI: 2003-450961/43.
DR P-FSDB; ADB65468.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide, or as a probe
CC is useful as a primer for synthesizing the polynucleotide, and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 1984 BP; 612 A; 357 C; 448 G; 567 T; 0 U; 0 Other;
XX
Query Match 72.0%; Score 1601; DB 9; Length 1984;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1662; Conservative 0; Mismatches 5; Indels 48; Gaps 1;
```

```
Qy 510 ATATGACTTAAAGCCAGAGCAATGATGACAGCACTGCTGCTTTTGTCTAGAA 569
Db 117 ATATGACTTAAAGCCAGAGCAATGATGACAGCACTGCTGCTTTTGTCTAGAA 176
Qy 570 GACCACTATTAAAGTGCACAGGTATGATGATCTTCTAAAGCAGAGTGCAGTGA 629
Db 177 GACCACTATTAAAGTGCACAGGTATGATGATCTTCTAAAGCAGAGTGCAGTGA 236
Qy 630 GTGAACAGAGTGTGTTCAAAAGTTTAATGAAGAATATAACAGGCTTGAAAGATT 689
Db 237 GTGAACAGAGTGTGTTCAAAAGTTTAATGAAGAATATAACAGGCTTGAAAGATT 248
Qy 690 CTTGAAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 749
Db 249 CTTGAAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
Qy 750 ACATACCCCTCTCTCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809
Db 309 ACATACCCCTCTCTCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
Qy 810 TAAACTTCGACAAATACGACACTATTCTTGTGAGAGTGTGAGACATGACACACAG 869
Db 369 TAAACTTCGACAAATACGACACTATTCTTGTGAGAGTGTGAGACATGACACACAG 428
Qy 870 CAATTCGACTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
Db 429 CAATTCGACTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
Qy 930 CTGATTACTAAGCTTACTCTGTCAGAAATATACAGTATTCCTATCATATCTTGAGAG 989
```


CC molecules which correspond to the ESTs. Proteins encoded by the ESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA5926 to AAA45931 represent linker variants which are given in the exemplification of the present invention

CC XX Sequence 1006 BP; 314 A; 181 C; 251 G; 260 T; 0 U; 0 Other;

Query Match 39.2%; Score 871.4; DB 3; Length 1006;

Best Local Similarity 90.5%; Pred. No. 2,2e-227;

Matches 984; Conservative 0; Mismatches 1; Indels 102; Gaps 1;

```

QY 53 ATGCTTCTATATGCTCCATATTTGAGAGATAGTATTTCTGAGAAAATGAGACAT 112
DB 22 ATGCTTCTATATGCTCCATATTTGAGAGATAGTATTTCTGAGAAAATGAGACAT 81

QY 113 TTAAAGGAGCTCAGCTTCATCATCAGAAATGATGAGACCTTCTGAGAGATCAT 172
DB 82 TTAAAGGAGCTCAGCTTCATCATCAGAAATGATGAGACCTTCTGAGAGATCAT 141

QY 173 TTGATGAAAAGTGATTTGGCTCTGGGCGGAATTTTGGAAAACAGATGCTGTAGTGT 232
DB 142 TTGATGAAAAGTGATTTGGCTCTGGGCGGAATTTTGGAAAACAGATGCTGTAGTGT 201

QY 233 AATAAGGAGATATATATCCACAATGGTGTGTTTGGATGGAAAAGTTTGGAAAC 292
DB 202 AATAAGGAGATATATATCCACAATGGTGTGTTTGGATGGAAAAGTTTGGAAAC 261

QY 293 AGAGGTTTTCAAACAGCAGGTTTGAAGATGTGATAGCTCTGTTTCTGGAGAGATCT 352
DB 262 AGAGGTTTTCAAACAGCAGGTTTGAAGATGTGATAGCTCTGTTTCTGGAGAGATCT 321

QY 353 AGTATAGCTGCGAAGATATATCCACAACGGAACAGAGGTTTTCGAAGAGCGGCTAT 412
DB 322 AGTATAGCTGCGAAGATATATCCACAACGGAACAGAGGTTTTCGAAGAGCGGCTAT 376

QY 413 CGAATGGAATATATTCAGAAAGCTTCAGGGCATACAGAAAGGTGAGAGTATTTTC 472
DB 377 ----- 376

QY 473 CGAGGTGCGGTGAGAGATTTGCTCTAGGAAGTCCAAATATATGATGACCCAGAGCAA 532
DB 377 ----- AATAAGCTTATGACCCAGAGCAA 399

QY 533 TGTATGACGCGCATCTGGTGGCTTTTGGTCTTGAAGACAGATATTAAGGGACAGGT 592
DB 400 TGTATGACGCGCATCTGGTGGCTTTTGGTCTTGAAGACAGATATTAAGGGACAGGT 459

QY 400 TGTATGACGCGCATCTGGTGGCTTTTGGTCTTGAAGACAGATATTAAGGGACAGGT 459
DB 460 AATGATGATCTTCTCAAAAGCAGAAAGTGGCAGTGAAGTGAACAGGTGTTCAAAAGT 519

QY 593 AATGATGATCTTCTCAAAAGCAGAAAGTGGCAGTGAAGTGAACAGGTGTTCAAAAGT 652
DB 520 TTAATATGAAAGATATTAACAGGCTCTGGAAGAAATCTTGAAGTCAAGAGCAAGAGGA 579

QY 653 TTAATATGAAAGATATTAACAGGCTCTGGAAGAAATCTTGAAGTCAAGAGCAAGAGGA 712
DB 580 GGAAGAAATGATGATCTCAAGGACCAAAAGTGAACCTTACATACCCCTCTCCACTGAG 772

QY 713 GGAAGAAATGATGATCTCAAGGACCAAAAGTGAACCTTACATACCCCTCTCCACTGAG 772
DB 580 GGAAGAAATGATGATCTCAAGGACCAAAAGTGAACCTTACATACCCCTCTCCACTGAG 639

QY 773 GATGAGACTCCATCTTTGACATTTATCAGACAGGCAATTAACCTTGCACAAATACGACACT 832
DB 640 GATGAGACTCCATCTTTGACATTTATCAGACAGGCAATTAACCTTGCACAAATACGACACT 699

QY 833 ATTCTTGGAGAGTGTCTGACATGATCAGACAGCAATTCGACTTTTGAAGAGACT 892
DB 700 ATTCTTGGAGAGTGTCTGACATGATCAGACAGCAATTCGACTTTTGAAGAGACT 759

```

```

QY 893 AATCTGTGACAGACTGAAATACACATTTGCTTAAAGCTGTTATATAGCTTACTCT 952
DB 760 AATCTGTGACAGACTGAAATACACATTTGCTTAAAGCTGTTATATAGCTTACTCT 819

QY 953 GTGAAAATAATACATATTCATCATCATCTGACAGAGAGATTTGATGAGCTTGTCTCAA 1012
DB 820 GTGAAAATAATACATATTCATCATCATCTGACAGAGAGATTTGATGAGCTTGTCTCAA 879

QY 1013 ACAGGCTGTGGGAAGACTGCGGCTTTTCTCTACCAATTTTGGCTCATATGATCATGAT 1072
DB 880 ACAGGCTGTGGGAAGACTGCGGCTTTTCTCTACCAATTTTGGCTCATATGATCATGAT 939

QY 1073 GGAATTAATGCTCAGCTTTTAAAGATTTGAGAGAACAGAGTGTATTTATGAGACCA 1132
DB 940 GGAATTAATGCTCAGCTTTTAAAGATTTGAGAGAACAGAGTGTATTTATGAGACCA 999

QY 1133 ACTGAG 1139
DB 1000 ACTGAG 1006

RESULT 4
AAS40979
ID AAS40979 standard; cDNA; 677 BP.

QY AAS40979;
DB 17-DEC-2001 (first entry)

DE cDNA encoding novel human enzyme polypeptide #195.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001239.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218230P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225270P.

```

PR	14-AUG-2000	2000US-0225447P	PR	08-NOV-2000	2000US-0246532P
PR	14-AUG-2000	2000US-0225157P	PR	08-NOV-2000	2000US-0246609P
PR	14-AUG-2000	2000US-0225158P	PR	08-NOV-2000	2000US-0246610P
PR	14-AUG-2000	2000US-0225159P	PR	08-NOV-2000	2000US-0246611P
PR	22-AUG-2000	2000US-0226279P	PR	17-NOV-2000	2000US-0246613P
PR	22-AUG-2000	2000US-0226681P	PR	17-NOV-2000	2000US-0246615P
PR	22-AUG-2000	2000US-0226686P	PR	17-NOV-2000	2000US-0246617P
PR	23-AUG-2000	2000US-0227182P	PR	17-NOV-2000	2000US-0246619P
PR	30-AUG-2000	2000US-0227009P	PR	17-NOV-2000	2000US-0246620P
PR	01-SEP-2000	2000US-0228924P	PR	17-NOV-2000	2000US-0249209P
PR	01-SEP-2000	2000US-0229287P	PR	17-NOV-2000	2000US-0249210P
PR	01-SEP-2000	2000US-0229343P	PR	17-NOV-2000	2000US-0249211P
PR	01-SEP-2000	2000US-0229344P	PR	17-NOV-2000	2000US-0249212P
PR	05-SEP-2000	2000US-0229345P	PR	17-NOV-2000	2000US-0249213P
PR	05-SEP-2000	2000US-0229509P	PR	17-NOV-2000	2000US-0249214P
PR	06-SEP-2000	2000US-0229513P	PR	17-NOV-2000	2000US-0249215P
PR	06-SEP-2000	2000US-0230437P	PR	17-NOV-2000	2000US-0249216P
PR	08-SEP-2000	2000US-0230438P	PR	17-NOV-2000	2000US-0249217P
PR	08-SEP-2000	2000US-0231242P	PR	17-NOV-2000	2000US-0249218P
PR	08-SEP-2000	2000US-0231243P	PR	17-NOV-2000	2000US-0249244P
PR	08-SEP-2000	2000US-0231244P	PR	17-NOV-2000	2000US-0249245P
PR	08-SEP-2000	2000US-0231411P	PR	17-NOV-2000	2000US-0249246P
PR	08-SEP-2000	2000US-0231414P	PR	17-NOV-2000	2000US-0249247P
PR	08-SEP-2000	2000US-0232080P	PR	17-NOV-2000	2000US-0249248P
PR	12-SEP-2000	2000US-0231081P	PR	17-NOV-2000	2000US-0249249P
PR	14-SEP-2000	2000US-0231968P	PR	01-DEC-2000	2000US-0250160P
PR	14-SEP-2000	2000US-0232397P	PR	01-DEC-2000	2000US-0250391P
PR	14-SEP-2000	2000US-0232398P	PR	05-DEC-2000	2000US-0251030P
PR	14-SEP-2000	2000US-0232399P	PR	05-DEC-2000	2000US-0251031P
PR	14-SEP-2000	2000US-0233400P	PR	05-DEC-2000	2000US-0251988P
PR	14-SEP-2000	2000US-0233401P	PR	06-DEC-2000	2000US-0256719P
PR	14-SEP-2000	2000US-0233063P	PR	06-DEC-2000	2000US-0251479P
PR	14-SEP-2000	2000US-0233064P	PR	08-DEC-2000	2000US-0251856P
PR	14-SEP-2000	2000US-0233065P	PR	08-DEC-2000	2000US-0251858P
PR	21-SEP-2000	2000US-0233223P	PR	08-DEC-2000	2000US-0251863P
PR	21-SEP-2000	2000US-0233227P	PR	08-DEC-2000	2000US-0251989P
PR	25-SEP-2000	2000US-0234997P	PR	11-DEC-2000	2000US-0254190P
PR	25-SEP-2000	2000US-0234998P	PR	05-JAN-2001	2001US-0259678P
PR	26-SEP-2000	2000US-0235484P	XX		
PR	27-SEP-2000	2000US-0235834P	XX		
PR	27-SEP-2000	2000US-0235836P	XX		
PR	29-SEP-2000	2000US-0236327P	XX		
PR	29-SEP-2000	2000US-0236367P	XX		
PR	29-SEP-2000	2000US-0236368P	XX		
PR	29-SEP-2000	2000US-0236369P	XX		
PR	29-SEP-2000	2000US-0236370P	XX		
PR	02-OCT-2000	2000US-0237037P	XX		
PR	02-OCT-2000	2000US-0237038P	XX		
PR	02-OCT-2000	2000US-0237039P	XX		
PR	02-OCT-2000	2000US-0237040P	XX		
PR	13-OCT-2000	2000US-0239935P	XX		
PR	13-OCT-2000	2000US-0239937P	XX		
PR	20-OCT-2000	2000US-0241809P	XX		

26-SEP-2002.
08-JAN-2002; 2002US-00044090.
28-JUL-2000; 2000US-0222469P.
08-JAN-2001; 2001US-0260483P.
(BAND/) BANDMAN O.
Bandman O;
WPI; 2003-110597/10.
Combination for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, comprises several cDNAs that are differentially expressed in activated vascular tissue.
Claim 1; Page; 18pp; English.
This invention relates to a combination comprising several cDNAs that are differentially expressed in activated vascular tissue. The invention also discloses a high throughput method for detecting differentially expressed cDNAs in a sample. The cDNAs of the invention may have
CC antiatherosclerotic; cyostatic; cardiant; hypotensive; antidiabetic;
CC synaerological; vasotropic and cerebroprotective activities and may be
CC used in gene therapy. The cDNAs of the invention may be used in a high-
CC throughput methods for detecting differential expression of one or more
CC cDNAs in a sample, or screening several molecules or compounds to
CC identify a molecule or compound that specifically binds a cDNA of the
CC invention. A protein encoded by the cDNA may be used to screen several
CC molecules or compounds to identify a ligand that specifically binds to
CC the protein, or to produce or purify a ligand that specifically binds to
CC be used to detect a protein in a sample or purify a natural or
CC recombinant protein from a sample. The nucleotides may be useful for
CC diagnosing, staging, treating, or monitoring the progression of treatment
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC genetic or gene expression analysis of several new nucleic acid
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http://seqdata.uspto.gov/sequence.html?docid=20020137081
XX
XX
SQ Sequence 2451 BP; 737 A; 436 C; 613 G; 665 T; 0 U; 0 Other;
Query Match 17.4%; Score 386.4; DB 7; Length 2451;
Best Local Similarity 58.7%; Pred No. 1.2e-94;
Matches 763; Conservative 0; Mismatches 501; Indels 36; Gaps 4;
QY 801 AACACGGCATTAATTCGACAAATACGACATATTCCTGGAGAGTGTGACATGATG 860
DB 503 AACACGGGGATTAATTCGAAATATGATGATATCCAGTAGAGCAACCGGACGTAAT 562
QY 861 CACCCACGACATTCGACTTTGAAGAAGCTAATCTGTGACACATGATAATACAA 920
DB 563 GTCTTCACATATTAAGAAATTTTACGATTTGAGATGAGGAGAAATTTATCATG 622
QY 921 TTGCTAAAGCTGTATATCTAGCTTACCTGTGCAAAATATACAGTATTCCTATCAT 980
DB 623 TTGAATTAATCTGCTATATCTGCTACTCCAGTGCAGAAAACATGCCATTTCTAT 682
QY 981 TTGACGAGACGATTTGATGCTTTGCTCAACAGGCTGTGGAGAACTGCGCTTTTC 1040
DB 683 AGGGAAAAGAGACTTAATGCTTGTGCCAAACAGGATCTGGGAAAATCTGCAATTTTC 742
QY 1041 TCTTACCAATTTTGGCTCA-----TATGATGATGATG 1073

DB 743 TTTTACCACTACTGATGATATATACAGATGCTCCAGGAAAGCTTTGAAGCTGTGA 802
QY 1074 GAATTAATGCGACATGCTTTTAAAGATTGACAGAACGAGTGTATTAATGTACACCA 1133
DB 803 AGGTAAATGGAAGATGAGCGCCGCAACATATCCATATCTTGGTTTATGCCCA 862
QY 1134 CTGAGAAATGGTCAACAGATTTATTTGAGAGCCAGAAATTTTCTTTTGGACTGTG 1193
DB 863 CAAAGAAATTTGGCTGTACAGATCTATGAGAGAGCCAGAAATTTTCTTCCATCCAT 922
QY 1194 TAAGAGCTGTGTATATATGAGGAGAAACCACTGTGGACATTCATTCGACAAATGAT 1253
DB 923 TTGCTCTTGTGTATGATTTATGATGTGTGTGATATGTGTCAGCAATTCGGGACT 982
QY 1254 AAGGTGTAAATATATATGATGCTTCTCTGTAACATGATGATATATATAGGCAAG 1313
DB 983 GTGATGCCATTTGTATGATGACCACTCAGAGACGTATGAGATATATGAAAGAGAA 1042
QY 1314 AGATGTGCTCAACACAGTCAATTAATTAATTTGATGATGATGATGATGATGATGAT 1373
DB 1043 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1102
QY 1374 TGGGTTTGTGTCGAAATGAAAGATTAATTTTCCAGAGATGATGATGATGATGATG 1433
DB 1103 TGGGATTTGAACCTCAGATATGCTGATGATGATGATGATGATGATGATGATGAT 1162
QY 1434 AGCCCAACCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1493
DB 1163 TTGCTACACAT 1222
QY 1494 AGTTTAAATGAT 1553
DB 1223 ACTTTTGTGATGAA--TATATCTTTTGTGCTGATGATGATGATGATGATGATGAT 1279
QY 1554 ATGTTGACGAGACCGTCTCCAGTGTGATGATGATGATGATGATGATGATGATGATG 1613
DB 1280 ACATCAACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339
QY 1614 TTCTGCGAAACATAGG--GATGAAAGATGATGATGATGATGATGATGATGATGAT 1670
DB 1340 TTTAGGTGACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1399
QY 1671 CAGATTTTACTGCAATTTCTTTTCAAGAAATATCACTAATCAATGATGATGATGAT 1730
DB 1400 CAGATTTCTGAGAGATTTCTTATACATGAAAGATGATGATGATGATGATGATG 1459
QY 1731 ATCGGAAACAGAGAGACCGGAGCAAGCTTTGAGATTTTCTGTTGAAAGTCCAG 1790
DB 1460 ACCGTCACAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
QY 1791 TTCTGTGCTACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1850
DB 1520 TTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1579
QY 1851 TCAATTTGATCTTCTTCTACATGATGATGATGATGATGATGATGATGATGATGATG 1910
DB 1580 TCAATTTGATTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1639
QY 1911 GTTGTGGAAATCTGCGACAGAGATTTCTTTTGTGATGATGATGATGATGATGATG 1970
DB 1640 GTGTAGAAACCTGCGCTTGCACCTCATCTTTAA--TGAATAAATATGAAATAT 1696
QY 1971 CACAGCTCTAGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2030
DB 1697 CAAAGATTTGTTGATCTTTCTTGTGAGAGCTTAAACAAAGAAATGCTTCTTGT 1756
QY 2031 AAATTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2070
DB 1757 ATATGCTTATGAAACACACATCAAGAGGATGAGAGAGAGAGAGAGAGAGAGAG 1796

RESULT 7
AEX62951

RESULT 8
AAV69633 standard; DNA; 4416 BP.

XX AAV69633;
AC AAV69633;

DT 02-FEB-1999 (first entry)

DE Dead Box Y (DBY) gene sequence.

XX Non-recombining region; human; Y chromosome; X homologue; testis; DBY;
KM Infertility; sperm; gene alteration; inhibitor; Dead Box Y; ss.

XX Homo sapiens.

PN WO9846747-A2.

PD 22-OCT-1998.

PF 10-APR-1998; 98WO-US007115.

PR 11-APR-1997; 97US-0041877P.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Lahn BT, Page DC;

DR WPI, 1998-568729/48.

DR P-PSDB; AAW81503.

XX Novel genes in the non-combining region of Y chromosome - useful to
PT diagnose if male infertility or reduced sperm count has a genetic basis.

XX Claim 7; Fig 3A-B; 54pp; English.

XX This represents the nucleotide sequence of the Dead Box Y (DBY) gene. The
CC invention relates to genes occurring on the non-recombining region of the
CC human Y chromosome. The sequences fall into two classes: (1) X-homologous
CC DNA which are expressed in many organs, having functional X homologues
CC and (2) testis-specific DNA sequences. Y chromosomal DNA from males with
CC known conditions such as infertility and reduced sperm count can be
CC assessed using the invention to determine whether the condition is
CC associated with or caused by the occurrence of the gene or gene
CC alteration. Candidate inhibitors of the enzymatic activity of the genes
CC can be assessed using in vitro assays

SO Sequence 4416 BP; 1337 A; 740 C; 993 G; 1346 T; 0 U; 0 Other;

Query Match 17.2%; Score 383.2; DB 2; Length 4416;
Best Local Similarity 58.5%; Pred. No. 1.2e-93;
Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

QY 801 AGACAGGCGATTAACCTTCGACAAATGACACATATCTTGGAAGTCTGACATGATG 860
DB 529 ACACGGGKATTAACCTTCGAAATATGATATATCCAGTAGGCAACGGCAGTAATCT 588
QY 861 CACACACGACAAATTCGACTTTTGAAGAGCTATCTGTCCAGACTGATTAACA 920
DB 589 GTCTCCACATATGGAATTTAGCATATGACATGAGGAGAAATATCATGGGGAACA 648
QY 921 TTGCTAAAGCTGTATATCTAAGCTTACTCTCTGTGCAAAAATACAGTATCTTATCATAC 980
DB 649 TTGAACCTTACTGCTATCTGCTACTCCAGTCAAAAACATGCCATTCCTATATATA 708
QY 981 TTGCAGACGAGATTTGATGCTGTGCTCAACAGGGCTGGGAAGACTGGCTTTTC 1040
DB 709 AGGGAAGAAAGACTTAGGCTTGTGCCCAAAACAGATCTGGGAAAACCTGCACATTTTC 768
QY 1041 TCTTACCAATTTTGCTCA-----TATGATGATGATG 1073
DB 769 TTTTACCATATCTGATGATATATACAGATGCTCCAGAGAGACCTTTGAAGGCTGTGA 828
QY 1074 GATATACGCGAGTGTGTTTAAAGAGTTGACGAGAACGAGATGATTTATGTAGACCA 1133

DB 829 AGAAAAATGGAAGGTATGGGGCCGCAACAAATATCCAAATATCTTGTTTGGCCCA 888
QY 1134 CTCGAAATTTGCTAACCAGATTTATTTGGAACCGAATAATTTCTTTGGACTTGTG 1193
DB 889 CAAAGAAATTTGGCTGATCAAGATCTATGAGGAGACGAAAAATTTTCTTACCGATCTAGAG 948
QY 1194 TAAAGCTGTGTATATATGAGGGAACCCAGCTGGGACATTAATTCGACAAATATGATAC 1253
DB 949 TTGCTCTGTGTATGTTATGTTGTGTCTATATTTGGTTCAGCAGATTCGGGACTTGTAAC 1008
QY 1254 AAGCTGTATATATTTATGCTTACTCTGGAAGACTGATGATATCATGCGAAGAA 1313
DB 1009 GTGATGCCACTTGTATGATGACCACTCCAGAGGCTGATGATATGATGAAAGGAA 1068
QY 1314 AGATTGCTCAACACAGATCAATTAATCTAGTTTGTGATGAGTGCATGCGATGATGATA 1373
DB 1069 AGATTGATTTAGACTTGTGCAAGTACTTATGTTGATGATGAGCTGATAGATCTGATA 1128
QY 1374 TGGGTTTGTCCAGAAATGAAAGTTAAATTTCTGCCAGGAATGCCATCAAGGAAC 1433
DB 1129 TGGGATTTGAACCTCAGATACGTCTGATATAGTTGAACAAGATACTATGCCCAAGGCG 1188
QY 1434 AGGCGCAACCCCTTATGTTAGTGAAGTCACTTTTCCAGAGGAATTCAGAGTTGGCTGCGAG 1493
DB 1189 TTGCTCACACATGATGTTTATGCTTACTTTCTTAAGAAATATACAGATGCTTGTCTGTG 1248
QY 1494 AGTTTAAATCAAAATTAATCTGTTTGTGCTGTGGAACAAGTGGGTGAGACATGATGAG 1553
DB 1249 ACTTTTGTATGAA---TATATCTTTTGGCTGTAGGACAGATGAGCTCTACTCTGAGA 1305
QY 1554 ATGTTCAGCAGACCGTTCTCCAAAGTTGGCCAGTTCTCAAAAAGAAAAGCTGTGAAA 1613
DB 1306 ACATTCACACAGAAAGTATGTTGGGTGAAAGACTTAATAAACGTCATTTCTACATGGA 1365
QY 1614 TTCTGCAAAACATAGG---GATGAAAGAACTATGCTTTGTGAACTAAGAAAAG 1670
DB 1366 TTTTATGTTGACAGAGGAGATCTACTTACTTACTTGTGTGAGACCAAAAAGGAG 1425
QY 1671 CAGATTTTACTGCACTTTTCTTTGTGCAAAAAAATATCACTACATCAAGTATCATGTG 1730
DB 1426 CAGATTTCCCTGGAGATTTCTTATACATGAAGATATGCTGTACTATGATTTCAATGAG 1485
QY 1731 ATCGGGAACGAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTGGAAAGTGGCCAG 1790
DB 1486 ACCGTTCAAGAGATTCAGAGAGAGCCCTTACCAAGTTTCTCTCAGGAAAAGCCCA 1545
QY 1791 TTCTTGTCTACTTCAATGACTGTCAGAGGGCTGATATGAAATGTGCAACATGTTA 1850
DB 1546 TTCTAGGCTTACAGCTGTGTCAGCAGAGACTATGACATTTCAATGTGAGCAGATGTTA 1605
QY 1851 TCAATTTTGAATCTTCTTCTTACCATGATGATATGTTGATGAAATGGGCTGATGCTGTC 1910
DB 1606 TCAATTTTGAATTTTCCAAAGTATGATATGAAATATGATGATGATGATGATGATGATGAT 1665
QY 1911 GTTGGGAATTAATGCAAGCAATTTCTTTTGTGATCTTGATTCGATTAACCAATTTAG 1970
DB 1666 GTGAGGAACCTGAGGCTTGGCACTCACTTCTTAA---TGAATAAATATGAATATATA 1722
QY 1971 CACAGCTCTAGTAAAGATATGACAGATGCTCAACAGATGTTCTGTGATGTTGGAG 2030
DB 1723 CAAAGATTTGTGATCTTCTTGTGAAAGCTTAAATAAAGTGCCTTCTTGTGGGAAA 1782
QY 2031 AAATGCTTTAGTATACATATCTCTGCTGATGATGATG 2070
DB 1783 ATATGCTTATGAAACACCACTACAAAGGTGCGAGTGTG 1822

RESULT 9
ACA64934 standard; DNA; 4416 BP.
XX ACA64934;
AC ACA64934;

XX 27-JUN-2003 (first entry)
 XX Human DBV DNA corresponding to AF000984.
 DE Human, chronic inflammatory joint disease; infection; tumour;
 XX antiinflammatory; cytostatic; antirheumatic;
 KM immunosuppressive; gene therapy; etiological pathogenicity; ds.
 OS Homo sapiens.
 XX DE10127572-A1.
 XX 05-DEC-2002.
 XX 30-MAY-2001; 2001DE-01027572.
 XX 30-MAY-2001; 2001DE-01027572.
 XX (PATH-) PATHOARRAY GMBH.
 XX Haepi T, Ungethem U, Blaess S;
 XX WPI; 2003-240797/24.
 XX Reagents for diagnosis, study and therapy of chronic inflammatory joint
 PT and other diseases, comprises any of many specified genes or derived
 PT proteins.
 XX Claim 1; Page; 12pp; German.
 XX This invention describes a novel reagent for diagnosis, molecular
 CC definition and therapy of chronic inflammatory joint diseases, and other
 CC inflammatory disorders, infective or tumour diseases in humans. The
 CC products of the invention have antiinflammatory, cyostatic,
 CC antirheumatic, antirheumatic and immunosuppressive activity and can be
 CC used for gene therapy. The reagent of the invention and any proteins and
 CC antibodies derived from it, are used (i) for analysing tissue and blood
 CC samples for medical diagnosis; (ii) for diagnosis and characterisation of
 CC chronic joint diseases, on the basis of molecular characterisation, and
 CC determining the etiological pathogenicity principle of as yet
 CC uncharacterised inflammatory diseases, also monitoring progression and/or
 CC treatment of disease, and optimisation of therapy and (iii) for
 CC developing treatments for inflammatory diseases, particularly of joints,
 CC infections and tumours. ACA64801-ACA64965 represent human polynucleotides
 CC used in the method of the invention
 XX Sequence 4416 BP; 1337 A; 740 C; 993 G; 1346 T; 0 U; 0 Other;
 SQ
 Query Match 17.2%; Score 383.2; DB 7; Length 4416;
 Best Local Similarity 58.5%; Pred.No. 1.2e-93;
 Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

QY 1074 GAATPACGCGAGTGGTTTAAAGATTGCGAGNACGAGATGATATTGTAGCCAA 1133
 DB 829 AGGAAATGGAAGGTATGGGGCGCCGAACAATATCCAAATCTGTTTATGCCCCAA 888
 QY 1134 CTGAGAAATGGCTACACACAGATTTATTTGGAGCCCAAAAATTTCTTTTGGACCTTGTG 1193
 DB 889 CAAGAGAAATGGCTGTACAGATCTATAGAGAGCCAAATAATTTCTTACCGATCTAGAG 948
 QY 1194 TAAGAGCTGTGTATATATATGAGGGAACCCAGCTGGACATTTCAATTCGACAAATAGTAC 1253
 DB 949 TTGCTCTTGTGTATATATATGAGGGAACCCAGCTGGACATTTCAATTCGACAAATAGTAC 1008
 QY 1254 AAGCTGTATATATATATATGAGGGAACCCAGCTGGACATTTCAATTCGACAAATAGTAC 1313
 DB 1009 GTGATGCGCACTTGTATAGTACCACTCAAGGACGCTATGATATATGAGGGAACGGA 1068
 QY 1314 AGATTTGCTCAAAACAGATCTAATTAATTTTGGATGGAAGCTGATGCTGATGATA 1373
 DB 1069 AGATTTGCTCAAAACAGATCTAATTAATTTTGGATGGAAGCTGATGCTGATGATA 1128
 QY 1374 TGGCTTTGCTCAAAATGAGAAATGATTTTCTTCCAGGAATGCCATCAAGGAAC 1433
 DB 1129 TGGATTTGAACTTCAGATACGCTGATAGTGAACAAGATTAATGACCAAGGCGC 1188
 QY 1434 AGCGCAAAACCTTATGCTTCAAGTGCACATTTTCCAGGAAATTCAGAGTGGCTGAG 1493
 DB 1189 TTGTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1248
 QY 1494 AGTTTAAAGCAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1553
 DB 1249 ACTTTTGAAGAA---TATATCTTTTGGCTGATGAGAGATGAGCTTACCTGAGA 1305
 QY 1554 ATGTTCAACAGACCGCTTCCAGATGTCACATTTCTCAAAAAGAGAAAGCTGCTGAA 1613
 DB 1306 ACATCACACAGAAATGATTTGGTGGAGAACTTATGATTAACGATCTTCTACGAGCA 1365
 QY 1614 TTCTCGGAAACCTAGG---GATGAAGAATGATGATGATGATGATGATGATGATGATG 1670
 DB 1366 TTTTGGGCAACAGAGAGTGTATCTTACTTATGATGATGATGATGATGATGATGATG 1425
 QY 1671 CAGATTTACTGCAACTTTCTTGTGCAAAAATAATCACTACAAATATCCATGCTG 1730
 DB 1426 CAGATTTCTCGAGATTTCTTATATCAATGAAGATGATGATGATGATGATGATGATG 1485
 QY 1731 ATGCGAACAAGAGAGAGGAGCAAGCTTTGGAGATTTTGGTGGAAAGTCCAG 1790
 DB 1486 ACCGGTCACAGAGATGAGAGAGAGGAGGCTTACACAGTTTCTGACAGAAAAGCCCA 1545
 QY 1791 TTCTTGTGCTACTTCAAGTGTGCGAGAGGCTGATATGGAATGCAACATGTTA 1850
 DB 1546 TTCTTGTGCTACTTCAAGTGTGCGAGAGGCTGATATGGAATGCAACATGTTA 1605
 QY 1851 TCAATTTGATCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1910
 DB 1606 TCAATTTGATCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1665
 QY 1911 GTTGGGGAATCTGCGACAGAAATTTCTTTTGTGATCTTGAATCGATTAACATTTAG 1970
 DB 1666 GTTGGGGAATCTGCGACAGAAATTTCTTTTGTGATCTTGAATCGATTAACATTTAG 1722
 QY 1971 CACAGCTCTAGTAAAGATTAAGATGATGATGATGATGATGATGATGATGATGATGATG 2030
 DB 1723 CAAAGATTTGTTGATCTTCTTGAAGAGCTTAACAAAGATGCTTCTTGTGGAAA 1782
 QY 2031 AAATTCCTTTAGTATATATATCTTCCGCTGAGAGGTAG 2070
 DB 1783 ATATGCTTATGAACCACTACAAAGGTGAGTGTGAG 1822

RESULT 10
 ADD15196
 ID ADD15196 standard; DNA; 4416 BP.
 XX

Db 1783 ATATGGCTTATGAAACACCACTACAGGAGTGGAGTCTGTGG 1822

RESULT 11
ABI99243
ID ABI99243 standard; cDNA; 2187 BP.
XX
AC ABI99243;
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:67.
XX
KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN M0200188188-A2.
XX
PD 22-NOV-2001.
XX
PE 18-MAY-2001; 2001MO-JP004192.
XX
PR 18-MAY-2000; 2000JP-00145977.
XX
PA (UNINT-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR WPI; 2002-034733/04.
XX
PT P-PSDB; ABB57045.
XX
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or by
PT determining the expression profile of a gene group comprising these
PT genes.
XX
PS Claim 2; Page 210-215; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
CC protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or
CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR
CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention
XX
SQ Sequence 2187 BP; 635 A; 409 C; 596 G; 547 T; 0 U; 0 Other:
Query Match 16.7%; Score 371; DB 6; Length 2187;
Best Local Similarity 57.7%; Pred.No.1.8e-90;
Matches 759; Conservative 0; Mismatches 520; Indels 36; Gaps 4;

QY 786 TCTTTGCATTAACAGACAGGATTAAGTTCGCAATAGACATCTTGTGGAG 845
Db 472 TCTTTTCGAGGCAATCTGGGATTAACCTTGGAATATGAGCATTCGATCGAG 531
QY 846 TGTCTGCATGATGACACACAGCAATTCGACTTTGAGAAGCTAATCTGTGAGA 905
Db 532 CAACAGGAGCAACAGCTCTCCACACATGAAAGTTTCAGTATGAGATGGAGAA 591
QY 906 CACTGAATTAACAACTTCTAAGTGTGTTTACTTAAGCTTAATCCCTGTGAAATACA 965
Db 592 TTATTATGGAGAACTTGAAGCTTACTGTTATACCTCGCCCAATCCCACTGAGAGCATG 651

QY 966 GTATTCCTATCACTACTGACAGAGATTTGATGAGCTGTGCTCAAAACAGGCTGCGGA 1025
Db 652 CTATTCCTATTAATCAAGAGAGAAAGACCTTGATGAGCTGTGCTCAAAACAGGCTGCGGA 711
QY 1026 AGACTGGGCTTTTCTCTCAACAAATTTGGCTCATATGATGATGATG----- 1074
Db 712 AAACGACAGACTTCTCTTGGCCATCTTGAGTCAATCTATGCTGATGCTCCAGAGAG 771
QY 1075 -----AATACTGCCAGTGGTTTAAAGATTGCCAGAACCAAGTGA 1118
Db 772 CTGAGAGGCTATGAGAGAAATGAAATATGAGCCGCTGCAACATGATCAATCTCTT 831
QY 1119 TTATGTAGCAACCACTGAGAAATGTGCAACGATTTATTTGAAAGCCAGAAATTTT 1178
Db 832 TGTACTGCGACCAACGAGAGAAATTTGCGAGATCTATGAGAAAGCCAGAAATTTCT 891
QY 1179 CTTTGGGACTTGTGTAGAGCTGTGTATATATGAGGGGAAACCAAGTGGGACATTCAA 1238
Db 892 CATACCGATCTAGAGTCCGTCCTTGGCTTTATGAGTGTGCTGAATTTGGCAGCA 951
QY 1239 TTCCGCAATATGTAACAAGGCTGTATATATATATGCTACTCTGGAAGACTGATGATA 1298
Db 952 TTCCGAGACTTAGAAGCTGATGCCACTGTATAGTACCACTCCAGAGAGCTTATGATA 1011
QY 1299 TCATAGGCAAGAAAGATTTGCTCAACAGATCAATATCTTGTGTTGATGATGAGCTG 1358
Db 1012 TGATGAGAGAGAGGAGATCGGGTTAGACTTCTGCAATACCTGCTGTATGATGAGCTG 1071
QY 1359 ATGCGATGTGGATATATGAGTGTGTTGGTCCAGAAATGAAAGATTAATTTCTGCCAGAA 1418
Db 1072 ACCGATGTTGATATGAGGTTTGAACCTCGATACGAAAGATGTTCAACAAGACATA 1131
QY 1419 TCCATCAAGAGAACAGGCGCAAAACCTTATGTTCACTGCAACTTTTCCAGAGAAATTC 1478
Db 1132 TGCCTCAAAAGATGTCCGCCACACTATGATGTTAGTACTTTCTTCAAGGAAATAC 1191
QY 1479 AAGGTTGCTGCGAAGTTTAAAGTCAATATCTGTTGCTGTGAGACAGTGG 1538
Db 1192 AGATCTGCGCCGTATTTCTTAGTG--AGTACATATTTCTGTGCTGATGAGAGATTTG 1248
QY 1539 GTGAGCATGTAGATGTTGACAGACCGTCTTCCAAAGTTGCGCAATCTCAAAAGAG 1598
Db 1249 GGTCTACTTGAGAGAACTACACAAAGAGTGTGCTGAGAGATATAGCAAAAGCT 1308
QY 1599 AAAAGCTGTTGAATTTCTGCAAAACATAG--GATGAAAGACTATGCTTTGTG 1655
Db 1309 CATTTCTGTTGACCTTCTTAATGCAACAGCAAGAAATTCCTGACTGATGTTGTGG 1368
QY 1656 AAAGTAAAGAAAGACAGATTTTACTGCAACTTTCTTGTGCAAGAAATATCACTA 1715
Db 1369 AATCCAAAGAGGCGAGATTTGCTGAGAGATTTCTTATACATGAAAGATATGCTTGT 1428
QY 1716 CAAGTATTCATGATGATGGGAAACAGAGAGCGGAGCAAGCTTTGGAGATTTGCT 1775
Db 1429 CCAGTATCCATGAGACCGTTCTCAGAGAGATGAGAAAGAGCCCTTCAACAGTTC 1488
QY 1776 TTGGAAGTGGCCAGTCTGTGCTACTTCACTGATGAGTCCAGAGAGGCTGATTTGAA 1835
Db 1489 CAGGAAAGAGCCCAATTTAGTGGCTTACAGAGTACAGAGAGAGAGATGATTTTCAA 1548
QY 1836 ATGTGCAATGTTATCAATTTTGAATCTTCTTCAACATGATGATGATATGTTCA 1895
Db 1549 ATGTGAGCATGTTATATATTTTGAATGCTAGATGATGATGATGATGATGATGAT 1608
QY 1896 TTGGGAGTCTGTGTTGTGGGAATCTGAGAGCAATTTCTTTTGTGTTGAT 1955
Db 1609 TAGGCGGATCAGGCGGTGTGGAAACCTTGTGCTTGCACCTCAATCTTTAA--TGAA 1665
QY 1956 CGGATTAACATTTAGACAGAGCTCTAGTAAAGATATGACAGATGCTCAACAGATGTC 2015
Db 1666 GGAATTAATATCAAAAGATTTTACTGATCTTCTTGTGAGGAAAGAGAGTGC 1725
QY 2016 CTGCAATGTTGAGAAAGATTTGCTTTAGTACATATCTCTGCTTCACTGATG 2070

Db 1726 CTTCTTGGTTAGAGAACATGCTTTTGAACACCACTACAGGAGTACAGTCTGG 1780

RESULT 12
ACD13409
ID ACD13409 standard; cDNA; 3158 BP.
XX
XX ACD13409;
XX
XX 13-AUG-2003 (first entry)
XX
DE Human DNA encoding a p53 modifier, SEQ ID 81.
XX
XX Human, ss; gene; p53 modifier; cytosolic; cancer; cytosolic;
XX antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;
XX kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX apoptotic disorder; cell proliferation disorder.
XX Homo sapiens.
XX
XX W0200299122-A1.
XX
XX 12-DEC-2002.
XX
XX 03-JUN-2002; 2002MO-US017382.
XX
XX 05-JUN-2001; 2001US-0296076P.
XX 10-OCT-2001; 2001US-0328605P.
XX 15-FEB-2002; 2002US-0357253P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Liang H, Li D, Funke RP,
XX WPI, 2003-156859/15.
XX P-PDB; ABO07236.
XX
XX Identifying modulators of the p53 pathway for use in treating apoptotic
XX or cell proliferation disorders, comprises screening for agents that
XX modulate activity of a human ortholog of genes that modify the p53
XX pathway in Drosophila.
XX
XX Example 2; Page 363-364; 678bp; English.
XX
XX The invention relates to identifying (M1) a candidate p53 pathway
XX modulating agent, by contacting an assay system comprising a purified HM
XX polypeptide (human orthologue of genes that modify the p53 pathway in
XX Drosophila) or nucleic acid with a test agent under conditions, where but
XX for the presence of the test agent, the system provides a reference
XX activity, and detecting a test agent-biased activity of the assay system.
XX Also included are modulating (M2) a p53 pathway of a cell (comprising
XX contacting a cell defective in p53 function with a candidate modulator
XX that specifically binds to a HM polypeptide comprising an HM amino acid
XX sequence, where p53 function is restored), modulating (M3) a p53 pathway
XX in a mammalian cell (comprising contacting the cell with an agent that
XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
XX from the patient; (b) contacting the sample with a probe for HM
XX expression; (c) comparing the results with a control; and (d) determining
XX whether the comparison indicates a likelihood disease). (M1) is useful
XX for identifying modulators of the p53 pathway. A probe for HM expression
XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
XX in a patient, where the cancer has greater than 25 % expression level.
XX Modulators identified by (M1) are useful in a variety of diagnostic and
XX therapeutic applications, where disease or disorder prognosis is related
XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
XX proliferation disorders (e.g. cancer). Another two new methods (M2 and
XX M3) are useful for modulating the p53 pathway of a cell, thus restoring
XX the p53 function of the cell, so that the cell undergoes normal
XX proliferation or progression through the cell cycle. (M2) and (M3) are
XX also useful for treating defects in the p53 pathway such as angiogenic,
XX apoptotic or cell proliferation disorders. The present sequence is an HM

CC nucleic acid encoding a p53 pathway modifying protein
XX
XX Sequence 3158 BP; 906 A; 607 C; 814 G; 831 T; 0 U; 0 Other;
SQ
Query Match 16.7%; Score 371; DB 7; Length 3158;
Best Local Similarity 57.7%; Pred. No. 2.2e-90;
Matches 759; Conservative 0; Mismatches 520; Indels 36; Gaps 4;
QY TCTTTGACATTATCAGACAGGACCTTGAACAAATGACACATTTCTTGTGAG 845
DB TCTTTTCTGAGGACACACCTGGATTATTTTGGAAATACGATGATTCAGTTGAG 817
QY TGTCTGACATGATGACACACAGCAATTCGATTTTGAAGAAGCTAATCTGTGACA 905
DB CACAGGACCAACATGCTCTCCACATATGAAAGTTTCAGTATGTTGAGATGGAGAAA 877
QY CACGAAATPACAAATGCTTAAAGCTGTATTAAGCTTACTCTCTGCAAAAATACA 955
DB TTATCATGGGAAACATTTAGCTTACTGTTATCTGCCCACTCCAGTGCAGAAAGCATG 937
QY GTATTCCTATCATCTTGCAGAGACGAGATTGATGCTGTGCTCAACAGGCTTGGA 1025
DB CTATTCCTATTTATCAAGAGAAAAGAGACTGATGCTGTGCTCAACAGGCTTGGA 997
QY AGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATG----- 1074
DB AAACCTGACGATTTCTGTTTCCCATCTTGATGATGATTTATTCAGATGCTGCAGAG 1057
QY 1026 AGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATG----- 1074
DB 998 AAACCTGACGATTTCTGTTTCCCATCTTGATGATGATTTATTCAGATGCTGCAGAG 1057
QY 1075 -----AATACCTGCAAGTGTGTTTAAAGTTGACGAGAACGAGATGTA 1118
DB 1058 CTTGAGGCGCATGAAGAAATGAAAGATATGGGCGCCGCAACATATCCATCTCT 1117
QY 1119 TTATTTGACACCACTGCAAAATTTGCTCAACCAATTTATTTTGAAGCCAGAAAATTT 1178
DB 1118 TGGTATTTAGACCAACAGAGAGTTGGCAGTACAGATCTACAGAGAGCAGAAAATTTT 1177
QY 1179 CTTTGGGACCTTGTATGAGAGCTGTGTTATATATATGAGGGAACCCAGCTGGACATTTCA 1238
DB 1178 CATACGATTTAAGTTTCTGCTGCTGTTTATATGTTGCTGCTGCTGCTGCTGCTGCTG 1237
QY 1239 TTGACAAATATGACAGGCTGTAATATATATATATATATATATATATATATATATAT 1298
DB 1238 TTGACAGCTTGAACCTGATGTCATTTGTTAGTACCACTCCAGAGCCTATGATGATTA 1297
QY 1299 TCATAGGCAAGAAAGATTTGCTCAACCAATCAATCAATCAATCAATCAATCAATCAAT 1358
DB 1298 TGAATGAAAGAGAAAGATTTGATGACTTTGCAAAATATGTTGATGATGATGATGATG 1357
QY 1359 ATGCAATGTTGATATGATGTTTGTGCAAAATGAAAGATTAATTTCTGCCAGAA 1418
DB 1358 ATGCAATGTTGATATGATGTTTGTGCAAAATGAAAGATTAATTTCTGCCAGAA 1417
QY 1419 TGGCATCAAGAAAGAGAGGCGCAACCTTATGTTCAAGTCAACTTTCCAGAGAAATTC 1478
DB 1418 TGGCTCAAGAGGTTGCGCCACACATATATATATATATATATATATATATATATAT 1477
QY 1479 AAAGTTGCTGACAGATTTTAAAGTCAATATCTGTTGTGCTGTGACAAAGTGG 1538
DB 1478 AAGATGCTGCTGCTGATTTCTTATGATGA---TATATCTTGTGCTGTGACAAAGTGG 1534
QY 1539 GTGAGCATATGAGATGTTCAAGACCCCTTCTCCAAAGTTGCCAATTTCCAAAAGAG 1598
DB 1535 GCTCTACCTCTGAAACATCAACAGAAAGTATGTTGGGTGGAAGAAATCCAGCAACGAT 1594
QY 1599 AAAGTCTGTTGAATTTGCGCAACATATGAG---GAGTGAAGAAATATGCTTGTGTTG 1655
DB 1595 CATTTTGTGATCTCTTAAATGCAACGCAAGATTCATGACCTTATGTTGTTGG 1654
QY 1656 AAATGAAGAAAGAGATTTTACTGCAACTTTTCTTGTCAAGAAATAATCACTA 1715
DB 1655 AAGCCAAAAGGATGCAATTTCTGAGAGATTTCTTATACATGAAGATACGATGTA 1714
QY 1716 CAAGTATCATGCTATGCGGAAACAGAGAGCGGAGCAACGCTTGTGAATTTTCCT 1775

DB 1715 CCAGCATCCATGAGACCCCTTCTCAGAGGATAGAGAAAGAGCCCTTCAACCACTTCGGCT 1774
 QY 1776 TTGAAAAGTCCCACTTCTTGTGTAATTAGTCAAGTCCAGAGGCTGATATTGAAA 1835
 DB 1775 CAGGAAAAAGCCCAATTTAGTGGCTACAGCAGTAGCAGAAAGAGAGCTGACATTTTCA 1834
 QY 1836 AGTGCACACATGTTATTCATTTTGTATCTTCTTACCATGATGATATGTTATGAAA 1895
 DB 1835 AGTGAACATGTTATTCATTTTGTATCTTCTTACCATGATGATATGTTATGAAA 1894
 QY 1896 TTGGGGCTGATCTGCTGTTGTGGAATACCTGCAAGCAATTCCTTTTGTATGTTAAT 1955
 DB 1895 TTGGGCTGATCGGAGGTGAGGAACTTGGCTGGCAACCTCAATCTTTAA--CGAGA 1951
 QY 1956 CGGATACCATTTAGCAGAGCCTCTAGTAAAGATTGACAGATGCTCAAGAGATGTC 2015
 DB 1952 GGAACATTAATTTACTTAAGATTTGTTGATCTTCTTGTGTAAGCTAAACAGAAAGTGC 2011
 QY 2016 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCAAGTGTAG 2070
 DB 2012 CCTCTTGTGTTAGAAAACATGCTTATGACACACACTCAAGGGTAGCAGTGTGG 2066

RESULT 13
 AAV69632 standard; DNA, 3408 BP.
 ID AAV69632
 AC AAV69632;
 DT 02-FEB-1999 (first entry)
 DE Dead Box X (DBX) gene short transcript nucleic acid sequence.
 XX Non-recombining region; human; Y chromosome; X homologue; testis; DBX;
 KM infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
 OS Homo sapiens.
 XX MO9846747-A2.
 PN 22-OCT-1998.
 PD 10-APR-1998; 98WO-US007115.
 PF 11-APR-1997; 97US-0041877P.
 PR (MHED) WHITEHEAD INST BIOMEDICAL RES.
 XX Lahn BT, Page DC;
 PI WPI; 1998-568729/48.
 DR P-PSDB; AAW81502.
 DR Novel genes in the non-recombining region of Y chromosome - useful to
 PT diagnose if male infertility or reduced sperm count has a genetic basis.
 PS Disclosure; Fig 3A-B; Sapp; English.
 XX This represents the nucleotide sequence of the Dead Box X (DBX) gene
 CC short transcript. The invention relates to genes occurring on the non-
 CC recombining region of the human Y chromosome. The sequences fall into two
 CC classes: (1) X-homologous DNA which are expressed in many organs, having
 CC functional X homologues and (2) testis-specific DNA sequences. Y
 CC chromosomal DNA from males with known conditions such as infertility and
 CC reduced sperm count can be assessed using the invention to determine
 CC whether the condition is associated with or caused by the occurrence of
 CC the gene or gene alteration. Candidate inhibitors of the enzymatic
 CC activity of the genes can be assessed using in vitro assays

SQ Sequence 3408 BP; 951 A; 734 C; 923 G; 800 T; 0 U; 0 Other;
 Query Match 16.6%; Score 369.4; DB 2; Length 3408;

Best Local Similarity 57.6%; Pred. No. 6.1e-90;
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;
 QY 786 TCTTTGCACTTATCAGACAGCATTAATCTTCGAAATTCACACTATTTCTTGTGAA 845
 DB 1305 TCTTTTCTGAGGCACTGGGATTAATTTGAAATTCGATGACATTCACAGTTGAG 1364
 QY 846 TGCTGACATGATGACACACACATTCGATTTGAAAGAGTAATCTCTGTACGA 905
 DB 1365 CAACAGGCAACAACCTGCTTCCACATATGAAATTTCAATGATTTGAATGGAGAA 1424
 QY 906 CACTGATTAACAATGCTTAAAGCTGTTATTAATTAATCTTCTCTGCAAAATACA 965
 DB 1425 TTATCTGAGGAAATGAGCTTATCTGTTATCTCCCACTCCAGTCCAAAGCATG 1484
 QY 966 GATTCCTATCATCTTCCAGAGACGATTTGATGCTGTGCTCAACAGGCTTGGGA 1025
 DB 1485 CTATTCCTTATTAAGAAAGAAAGACATTTGATGCTTGTGCCAAACAGGCTTGGGA 1544
 QY 1026 AGACTGCGCTTTTCTCCATCAATTTGGCTATATGATCATGATG----- 1074
 DB 1545 AAACCTGACGATTTTCTGTTGCCATCTTGAATGATTTTCAAGTGGTCCAGGAGAG 1604
 QY 1075 -----AATACTGCGATGCTTTTAAAGTTGACAGAACGAGATGTA 1118
 DB 1605 CTTTGAAGGCCATGAAGAAATGAAAGTATGGGCGCCGCAACATACCAATCTCT 1664
 QY 1119 TTATTTAGACCAACTCGAGAAATTTGTCACACGATTTATTGGAAGCCAGAAATTTT 1178
 DB 1665 TGGTATTAGCACCAACGAGAGAGTTGGCAGTACAGATCTAGAAAGCCAGAAATTTT 1724
 QY 1179 CTTTGGGACTTGTGAAGAGCTGTGTATATATGAGGAGAACCCGCTGGGACATTCGA 1238
 DB 1725 CATACCATTTAGAGTTCCTTCTGCTGTTTATGATGAGTGGCATTTGTGTACAGCA 1784
 QY 1239 TTGCAAAATAGTACAAAGCTGTATATATATGATGCTACTCTGGAAGACTATGATA 1298
 DB 1785 TTGAGACTTGGAACTGAGATGCTCATTTGTATGAGCCATCCAGACGCTGTGTGATA 1844
 QY 1299 TCATGAGGAAAGAAAGATTTGGTCTCAACAGATCAATTAATAGTTTGGATGAAAGCTG 1358
 DB 1845 TGATGAAAGAGGAAAGATTTGATTTAGCTTTTGCAAAATCTTGTGTGTGATGAACTG 1904
 QY 1359 ATTCGATTTGATATGAGTTTGTGTGTCAGAAATGAAAGATTAATTTCTGCCCAGAA 1418
 DB 1905 ATTCGATTTGATATGAGTTTGTGTGTCAGATTCGTAATAATGTCGAACAAGATCTA 1964
 QY 1419 TGCCATCAAGAAACAGGCGCAACCCCTATATGTTCAATGCACTTTCCAGAGAAATTC 1478
 DB 1965 TGCTCCAAAGAGGTGTGCGCCACACTATGATTTTATGCTACTTTTCTTAAAGAAATAC 2024
 QY 1479 AAAGTTGGCTGCAAGATTTTAAAGTCAATTAATCTGTTGTGCTGTGCAAGATGG 1538
 DB 2025 AGATGCTGCTGCGATTTCTTAATGCAACAGGCAAGATATCACTGACCTTGTGTGG 2081
 QY 1539 GTGAGCATATGAGATGTTCAAGACGCTTCCCAAGTTGGCCGCTTCTCAAAAGAG 1598
 DB 2082 GCTTACTCTGAAATCAACACAGAAATGATTTGGGTGAGAAATTCAGCAAAAGCT 2141
 QY 1599 AAAGCTGTTGAAATCTGCAAACTAGG--GATGAAAGAACTATGCTTTGTG 1655
 DB 2142 CATTTCTGCTGACCTCTTAATGCAACAGGCAAGATATCACTGACCTTGTGTGG 2201
 QY 1656 AAACCTAAAGAAAGAGATTTTACTGCACTTTTCTTCAAGAAATTAATCACTA 1715
 DB 2202 AACCAAAAGAGGTCGACATTTCTTGAAGATTTCTTAACATGAAGGATACGATGTA 2261
 QY 1716 CAAGTATCATGATGATGAGAACAGAGAGAGGAGCAAGCTTTGAGATTTTGTCT 1775
 DB 2262 CCAGCATCATGAGACCGTTCTCAGAGGATGAGAAAGAGGCTTCAACAGTTCGCT 2321
 QY 1776 TTGAAATGCGCCAGTCTTGTGCTTCAATGATGCTGACAGAGGCTGATATTGAAA 1835

Db 2322 CAGAAAAAGCCCAATTTAGTGTCTACAGAGTAGACAGACAGAGACTGACATTTCAA 2381
 Qy 1836 ATGTCCACACAGCTTATCATTTTATCTTCTCTTACCATTTGATGATGTCAGAA 1895
 Db 2382 ATGTAAAACATGTTATCATTTTATCTTCTCTTACCATTTGATGATGTCAGAA 2441
 Qy 1836 TTGGGCGTATGCTGTGTGTGGGAATATCTGGCAGAGCAATTTCTTTTGTGATTTGAAT 1955
 Db 2442 TTGGGCGTATGCTGTGTGTGGGAATATCTGGCAGAGCAATTTCTTTTGTGATTTGAAT 2498
 Qy 1956 CGGATTAACCATTTTACGACAGCCTCTAGTAAAGTATGACAGATGCTCAACAGGATGTC 2015
 Db 2499 GGAACATTAATATTAATCTAAGATTTGTGATCTTTCTTTGTAAGCTAAACAGAGATGC 2558
 Qy 2016 CTGCATGCTGTAAGAAATGCTCTTATGATACATTCCTGCTTACGTTAGTGTA 2070
 Db 2559 CGTCTGCTTGAAGAAATGCTCTTATGATACATTCCTGCTTACGTTAGTGTA 2613
 RESULT 14
 AAA29207
 ID AAA29207 standard; DNA; 5321 BP.
 AC AAA29207;
 XX
 DT 12-SEP-2000 (first entry)
 DE Human DBX1 coding sequence.
 XX
 KM Topoisomerase III- α ; interacting protein; TIII- α IP; RNA helicase;
 KM DBX1; nucleosome; supercoiling; chromosome segregation; recombination;
 KM stabilisation; cell division; apoptosis; cell cycle regulation;
 KM cytotoxic; anti-tumour; DBX1; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 856..2844
 FT /tag= a
 FT /standard_name= "DBX1"
 FT /product= "RNA_helicase-like_protein"
 PN MO200032768-A1.
 PD 08-JUN-2000.
 XX
 PF 29-NOV-1999; 99WO-FR002952.
 XX
 PR 30-NOV-1998; 98FR-00015081.
 PA (AVET) AVENTIS PHARMA SA.
 XX
 PI Fournier A, Goulaouic H, Riou J;
 DR WPI; 2000-412316/35.
 DR F-PSDB; AAY96483.
 XX
 PT New nucleic acid encoding ligand for topoisomerase III α , useful for
 PT inhibiting the enzyme and in drug screening, e.g. for potential
 PT anticancer agents.
 XX
 PS Claim 6; Page 47-53; 68pp; French.
 CC This DNA encodes a DBX1, which shows homology with RNA helicases but the
 CC activity of a helicase has never been demonstrated and its function has
 CC not yet been identified. DBX1 possesses the 8 characteristic motifs of
 CC the helicases of the family "DEAD". In particular, it appears to be part
 CC of the sub-family represented by helicase P110. The DBX1 gene is situated
 CC on the X chromosome and its homologue, which is situated on the Y
 CC chromosome has 91 percent identity with a novel protein coding sequence.
 CC The novel protein (see AAY96482) is a human topoisomerase III- α IP
 CC interacting protein (TIII- α IP). TIII- α IP has structural features in
 CC common with RNA helicases (e.g. DBX1), which are involved in destruction

CC of the nucleosome structure, supercoiling of DNA, segregation of newly
 CC replicated chromosomes, and recombination and stabilisation of the
 CC genome. Agents that alter interaction between TIII- α IP and TIII- α
 CC therefore modulate cell division, replication, transcription,
 CC translation, splicing and DNA recombination or repair, so may slow cell
 CC growth, block the cell cycle or induce apoptosis. Antibodies and ligands
 CC of TIII- α IP are used to prevent, treat or alleviate diseases that
 CC involve abnormal regulation of the cell cycle, i.e. they are potential
 CC anti-tumour (cytotoxic) agents
 XX
 SQ Sequence 5321 BP; 1532 A; 1010 C; 1274 G; 1505 T; 0 U; 0 Other;
 Query Match 16.6%; Score 369.4; DB 3; Length 5321;
 Best Local Similarity 57.6%; Pred. No. 7.6e-90;
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;
 Qy 786 TCTTTGCAATTAATGACAGAGCAATTAATCTGCAAAATAGACACTTCTTGGAAG 845
 Db 1304 TCTTTTTCGAGGCAACCTGGATTAATTTGGAATAGAGACCTTCAATGAGG 1363
 Qy 846 TGTCTGACATGATGACACACACCAATTCGACTTTGAAAGACTAATCTGTGAGA 905
 Db 1364 CAACAGGCAACACTGCTCCACATATTAAGATTTCAGTATGTTGATGGAGAAA 1423
 Qy 906 CACTGAATTAACAATGCTTAAGCTGTTTAACTTAAGCTTACCTCTGTGAAAATACA 965
 Db 1424 TTATCATGGAAACATTGAGCTTAATCTGTTACTGCCCCAATCCAGTCAAAAGCATG 1483
 Qy 966 GTATTCCTATCATCTTGACAGACAGATTTGATGCTGTGCTCAACAGAGGCTGGA 1025
 Db 1484 CTATTCCTATTAATCAAGAAAGAAAGAACTGATGCTGTGCTCAACAGAGGCTGGA 1543
 Qy 1026 AGACTGGGGCTTTCTCTCTACCAATTTGGCTCATATGATGATGATG----- 1074
 Db 1544 AAACGTGACGATTTCTGTGCCCATTTGATGATGATGATGATGATGATGATGATGATG 1603
 Qy 1075 -----AATACTGCAGCTGTTTAAAGAGTTGACGGAACAGAGTGA 1118
 Db 1604 CTTGAGGGCCATGAAGAAATAGAAAGATAGAGATGAGGCCCCCAACATATCCATCTCT 1663
 Qy 1119 TTATTTAGACACCACTGCGAATTTGCTCAACAGATTTATTTGGAAGCCAGAAATTTT 1178
 Db 1664 TGGTATTAGACACCAACAGAGAGTTGGCAGTACAGATTTACGAAGAACAGAAATTTT 1723
 Qy 1179 CTTTGGGACTGTGTGAGAGCTGTGTATATATATGAGGGGAAACCCAGCTGGGCAATTC 1238
 Db 1724 CATACCATTTAGAGTTGCTGCTGCTGTGTATATGAGTGTGCGATATTTGTCACAGAGA 1783
 Qy 1239 TTGACAAATATGATCAAGGCTGTAATATATATGCTACTCTGGAAGACTGATGATA 1298
 Db 1784 TTGAGACTTGAACGCTGATGCAATTTGTTAGTACCACTCCAGAGAGCTAGTGGATA 1843
 Qy 1299 TCATAGGCAAAAGAAAGATTTGCTCAACAGATTCAAATCTTATTTGATGAAAGCTG 1358
 Db 1844 TGAATGAAGAGAAAGATTTGATTTGACTTTGCAAAATCTTGTGTGATGAAAGCTG 1903
 Qy 1359 ATGCAATGTTGATGATGATTTGCTCAAGAAATGAGAAATTTCTTCCACAGAA 1418
 Db 1904 ATGCAATGTTGATGATGATTTGAGGCTTGAAGCTTCAGATTTGATGATGATGATGAT 1963
 Qy 1419 TGGCATCAAGAAAGACAGGCCCCAATCTTATGTTAGTGAATTTTCCAGAGAAATTC 1478
 Db 1964 TGGCTCAAGAGGATGCTCCACACTATGATTTAGTCACTTTTCTTAAGGAAATAC 2023
 Qy 1479 AAAGTTGGCTCAGAGTTTAAAGTCAAAATATCTGTTGCTGTGGAACAAGTGG 1538
 Db 2024 AGATGCTGCTGTATTTCTTAAATGAA---TATATCTTCTTGGCTGATGAAAGTGG 2080
 Qy 1539 GTGAGCATGTAGAGATGTTCAAGACCTTCTCAAGTTGGCCAGTGTCTCAAAAGAG 1598
 Db 2081 GCTTACCTCTGAAACATCAACAGAAAGTGTGGTGAAGATCAACAAACGCT 2140
 Qy 1599 AAAGCTGCTGAAATTTGCGAAGCATAGG---CGATGAAGAACTATGCTTTGTTG 1655

Db 2141 CATTTCGCTTGAACCTCTTAATGCAACAGGAGGATTCAGTCACTTAGTGTTGG 2200
Qy 1656 AACTAGAAAAAGAGATTTTACTGCAACTTTCTTTCAGAAAAAATATCACTA 1715
Db 2201 AGACCAAAAAAGGTGCAAGATTCCTGGAGGATTTCTTATACATGAGATACGATGA 2260
Qy 1716 CAAGTATCATGATGATCGGGAACAGAGAGAGGAGGAGCACTCTTGGAGATTTGGT 1775
Db 2261 CAGCATCATGAGACCGTCTCAGAGGATAGAGAAAGGCCCTTACACAGTCCGCT 2320
Qy 1776 TTGAAAGTCCCAAGTTCTTTGCTACTTCACTAGCTCCAGAGGCTGGATTTGAA 1835
Db 2321 CAGAAAAAGCCCAATTTTATGCTACAGCACTACAGCAAGAGACTGACATTTCA 2380
Qy 1836 ATGTGCAACATGATATCAATTTTATCTCTCTCACTATGATGATGTTCAACGA 1895
Db 2381 ATGTGAAACATGTTATCAATTTTACTGCTGCAAGTATATGAAATATGATACATGA 2440
Qy 1896 TTGGGCGTACTGTCGTTGGGAATACCTGGAGCAAAATTTCTTTTATCTTGAAT 1955
Db 2441 TTGGTGGTACGGAGCGTGAAGAACTTGGCTGGCACTCATTCCTTTAA---CGAGA 2497
Qy 1956 CGGATACCATTTTACGACAGCTCTAGTAAAGTATGACAGATGCTCAACAGATGTC 2015
Db 2498 GGAACATAAATATTACTAAGGATTTGTTGATCTTCTTGAAGCTAAACAGAAATGC 2557
Qy 2016 CTGATGTTGGAAGAAATTCCTTATGATACATATTCCTGCTTCACTGTTAG 2070
Db 2558 CGTCTTGTTAGAAACATGCTTATGACACCACTAAGAGGTAGACAGTGG 2612

RESULT 15
AAV69631
ID AAV69631 standard; DNA; 5322 BP.
XX
AC AAV69631;
DT 02-FEB-1999 (first entry)
XX
DE Dead Box X (DBX) gene long transcript nucleic acid sequence.
XX
KM Non-recombining region; human; Y chromosome; X homologue; testis; DBX;
interfertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
XX
OS Homo sapiens.
XX
PN WO9846747-A2.
XX
PD 22-OCT-1998.
XX
PF 10-APR-1998; 98WO-US007115.
XX
PR 11-APR-1997; 97US-0041877P.
XX
PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Lahn BT, Page DC;
XX
DR WPI; 1998-568729/48.
XX
DR P-PSDB; AAM81501.
XX
PT Novel genes in the non-combining region of Y chromosome - useful to
diagnose if male infertility or reduced sperm count has a genetic basis.
XX
PS Disclosure; Fig 3A-B; Sapp; English.
XX
CC This represents the nucleotide sequence of the Dead Box X (DBX) gene long
transcript. The invention relates to genes occurring on the non-
recombining region of the human Y chromosome. The sequences fall into two
classes: (1) X-homologous DNA which are expressed in many organs, having
functional X homologues and (2) testis-specific DNA sequences. Y
chromosomal DNA from males with known conditions such as infertility and

CC reduced sperm count can be assessed using the invention to determine
CC whether the condition is associated with or caused by the occurrence of
CC the gene or gene alteration. Candidate inhibitors of the enzymatic
CC activity of the genes can be assessed using in vitro assays
XX
SQ Sequence 5322 BP; 1532 A; 1011 C; 1274 G; 1505 T; 0 U; 0 Other;

Query Match 16.6%; Score 369.4; DB 2; Length 5322;
Best Similarity 57.6%; Pred. No. 7.6e-90;
Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

Qy 786 TCTTTGACATTTATCAGACAGCATTAATCTGCAAAATAGCACATTTCTTGTGAAG 845
Db 1305 TCTTTTCGAGAGGAGACACTGGATTAATTTTGAATATGATGACATTCAGATTGAGG 1364
Qy 846 TGTCTGACATGATGACACACAGCAATTCTGACTTTGAAAGAGCTAATCTGTGAGA 905
Db 1365 CAACAGGCAACACTGTCTTCCACATATGAAAGTTTCACTGATGTGAGATGGAGAAA 1424
Qy 906 CACTGAATTAACAACTTGTCTTAAAGCTGTTATCTAAGCTTACTCTGTGCAAAATACA 965
Db 1425 TTATATGAGGAACATTTAGCTTACTGCTTATCTGCGCAACTCCAGTCAAAAGCATG 1484
Qy 966 GTATTCCTATCATACTTGACAGAGATTTGATGCTGTGCTCAACAGGGCTTGGGA 1025
Db 1485 CTATTCCTATTTATCAAGAGAAAAGAGACTTGATGGCTTGTGCCCAACAGGGCTTGGGA 1544
Qy 1026 AGACTGCGGCTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATGATGATG 1074
Db 1545 AACTGACAGCATTTGTGTCCTCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCT 1604
Qy 1075 -----AATTACTGCCAGTCTGTTTAAAGATTGCGAGAACCGAGTGA 1118
Db 1605 CTTTAGGGCCATGAGAAATATGAAAGTATGAGGCGCCGCAAAATATCACTCTCT 1664
Qy 1119 TTATTTGACCAACTGAGAAATTTGTCACACAGATTTATTTTGGAGCCAGAAATTTT 1178
Db 1665 TGGTATTTAGCAACCAAGAGAGTTGGGAGTACAGATCTTACAGAAACCGAGAAATTTT 1724
Qy 1179 CTTTGGGACTGTGTAAGAGCTGTTTATATATGAGGAGAACCCAGCTGGGACATTCAA 1238
Db 1725 CATACCGATCTAGATTTGCTTTCGCTTATATGATGATGATGATGATGATGATGATG 1784
Qy 1239 TTGCAAAATAGTACAGGCTGTAATATTTATTTGCTACTCTCTGGAAGCTGATGATTA 1298
Db 1785 TTGAGACTTGGAGAGCTGATGCCATTTGTTATGATGACCACTCCAGAGCTCTAGTGATA 1844
Qy 1299 TCATGGGCAAAAGAAATTTGGTCTCAACAGATCAAAATCTTATTTTGGATGAAGTGG 1358
Db 1845 TGAATGAAAGAGAAAGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1904
Qy 1359 ATGCAATTTGGATATGAGTTTGGTTCAGAAATGAAAGATTAATTTCTTCCAGGAA 1418
Db 1905 ATGCAATTTGGATATGAGTTTGGTTCAGAAATGAAAGATTAATTTCTTCCAGGAA 1964
Qy 1419 TGGCATCAAGAGAGAGCCCAACCTTATGTTGATGATGATGATGATGATGATGATGATG 1478
Db 1965 TGGCTCAAGAGGTGTCGCCCACTATGATTTGATGATGATGATGATGATGATGATGATG 2024
Qy 1479 AAGATTTGGCTGAGAGATTTTAAATCAATTTATCTGTTTGTGCTTGGACAGTGG 1538
Db 2025 AGATGCTGCTGCTGATTTCTTATGATGA---TATATCTTGTGGCTGATGAGAGTGG 2081
Qy 1539 GTGAGCATATGATATGATTTGACAGACCGCTTCTCAATTTGGCCAGTCTCAAAAAGG 1598
Db 2082 GCTCTACCTCTGAAACATCAACAGAAAGTATGTTGGTGGAGAGATACAGACAAAGCT 2141
Qy 1599 AAAAGCTCGTGAATTTCTGCGAAACATAG---GGATGAAGAACTATGCTTTTGTG 1655
Db 2142 CATTTCGCTTGAACCTCTTAATGCAACAGGAGATTCATGACCTTAGTGTGTTGG 2201
Qy 1656 AACTAGAAAAAGAGATTTTACTGCAACTTTTCTTGTCAAGAAAAATATCAACTA 1715

Db 2202 AGACCAAAAAGGTCAGATTCTCTGAGAGATTCTTATACATGAAGATACCATGTA 2261
QY 1716 CAAGTATCATGCTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTGCT 1775
Db 2262 CCAGCATCCATGAGACCGTTCTCAGAGGATAGAGAAAGGCCCTTCAACAGTCCGCT 2321
QY 1776 TTGAAAGTCCCGAGTCTTGCTCTTCACTTCACTGAGAGGCTGATATGAA 1835
Db 2322 CAGGAAAAAGCCCAATTTTATGCTTACAGCAGAGCAAGAGGACTGACATTTCAA 2381
QY 1836 ATGTGCAACATGTTATCAATTTTGATCTTCTTACCATTTGATGATATGTCAGAA 1895
Db 2382 ATGTGAAACATGTTATCAATTTTGACTTCCCAAGTGATTTGAAGATATGTACATGTA 2441
QY 1896 TTGGCGTACTGCTGTTGCGGAAATCTGGCAGAGCAATTTCTTTTGATCTTGAAT 1955
Db 2442 TTGCTCTTACGGAGCGTGTAGGAAACCTTGCGCACTCATTTCTTTAA---CGAGA 2498
QY 1956 CGGATTAACATTAGACACAGCCTCTAGTAAAGTATGACAGATGCTCAACAGATGTC 2015
Db 2499 GGAACATAAATATTACTTAAGATTTGTGATCTTCTTGTGAAGCTMAACAAGAGTGC 2558
QY 2016 CTGCATGCTGGAAGAAATGCTTTAGTACATACATCTCTGCTTCAAGTGTAG 2070
Db 2559 CGCTTGGTTAGAAAAACATGCTTTATGAACACCACTAACAGGCTACAGTCTGG 2613

Search completed: July 27, 2004, 06:51:00
Job time : 901 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 03:01:17 ; Search time 8597 Seconds

(without alignments)
11212.622 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224
Sequence: 1 actgaagtcacccatggggg.....aagtcctggcttcgatgca 2224

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6340544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2224	100.0	2224	AY004154	AY004154 Homo sapi
2	2214.4	99.6	2411	AF262962	AF262962 Homo sapi
3	1997.2	89.8	2363	BC047455	BC047455 Homo sapi
4	1939.6	87.2	2189	HSM802178	AL137462 Homo sapi
5	1601	72.0	1984	AX748127	AX748127 Sequence
6	1601	72.0	1984	AK093439	AK093439 Homo sapi
7	1594	71.7	3030	575275	S73275 RVG-vasa-1
8	1443	64.9	1930	MUSDVH	D14659 Mus musculu
9	751.2	33.8	1989	AB004836	AB004836 Gallus ga
10	712	32.0	2502	AF046043	AF046043 Xenopus l
11	686.8	30.9	2865	AB005147	AB005147 Danio rer
12	686.4	30.9	2492	DRY12007	Y12007 D. rerio vlg
13	651.4	29.3	2337	AF479820	AF479820 Cyprinus
14	619.4	27.9	2058	AB051835	AB051835 Oreochrom
15	619.4	27.9	2130	AB053467	AB053467 Oreochrom
16	612.4	27.5	2052	AB038252	AB038252 Leucopsar
17	612.2	27.5	2612	AB038256	AB038256 Oncorhyn
18	565	25.4	1512	AF479823	AF479823 Pantodon
19	552.8	24.9	2187	AB063484	AB063484 Oryzias l
20	548.2	24.6	1198	AF251800	AF251800 Danio dan
21	543.6	24.5	1730	AF479821	AF479821 Hypheasob
22	503.6	22.6	1586	AF520608	AF520608 Sparus au
23	503.4	22.6	1506	AF479824	AF479824 Melanctae
24	498.6	22.4	1816	AF479825	AF479825 Oncorhyn
25	493	22.2	504	BD110839	BD110839 EST and e
26	493	22.2	504	BD110839	BD110839 EST and e
27	487.4	21.9	2551	AB016603	AB016603 Clona int
28	487.4	21.9	2561	AB016604	AB016604 Clona int
29	457	20.5	1873	AB047385	AB047385 Ephydratia
30	456.6	20.5	1332	AB047383	AB047383 Hydrat mag
31	438.6	19.7	1764	AF510054	AF510054 Schistoc
32	438	19.7	3023	AB047803	AB047803 Clona sav
33	436.4	19.6	2777	AB047802	AB047802 Clona sav
34	434.6	19.5	2721	AB047802	AB047802 Hydrat mag
35	428.8	19.3	1449	AF479822	AF479822 Oryzias l
36	384.8	17.3	2713	BC034942	BC034942 Homo sapi
37	383.2	17.2	2319	HSARF000985	AF000985 Homo sapi
38	383.2	17.2	4416	HSARF000984	AF000984 Homo sapi
39	378.2	17.0	2384	AB047381	AB047381 Hydrat mag
40	377	17.0	524	AY100475	AY100475 Equus cab
41	371	16.7	2187	AX305316	AX305316 Sequence
42	371	16.7	2187	10 MUSRNABELI	L25126 Mus musculu
43	371	16.7	2201	AF061337	AF061337 Homo sapi
44	371	16.7	2499	BC011819	BC011819 Homo sapi
45	371	16.7	3158	HSU50553	US0553 Homo sapien

ALIGNMENTS

RESULT 1
AY004154
LOCUS AY004154 2224 bp mRNA linear PRI 25-AUG-2000
DEFINITION Homo sapiens DEAD box RNA helicase (VASA) mRNA, complete cds.
ACCESSION AY004154
VERSION AY004154.1 GI:9438226
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2224)
Castrillon,D.H., Quade,B.J., Wang,T.Y., Quigley,C. and Crum,C.P.
The human VASA gene is specifically expressed in the germ cell
lineage

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9585-9590 (2000)
 MEDLINE 20402578
 PUBMED 10920202
 REFERENCE 2 (bases 1 to 2224)
 AUTHORS Castriello, D.H.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-2000) Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA

FEATURES

source
 1..2224
 /organism="Homo sapiens"
 /mol_type="rRNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q11.2-12"
 /tissue_type="testis"
 /dev_stage="adult"
 1..2224
 /gene="VASA"
 14..2188
 /gene="VASA"
 /codon_start=1
 /product="dead box RNA helicase"
 /protein_id="AA06585.1"
 /db_xref="GI:9438227"
 /translation="MGDEWEAEINPMSSVVDIPEKDRYSGENEDNFRTPASSSEM
 DDGSRDRHFMKSGFASGRNFGNDAECNKRDNTSMGFGVGSFGNRSRFE
 DGGSGFWRSENDCEINPFRNRFSGRGYRDGNNSEAGPYRSGSGFGRGCGF
 GLGSDNDLDDECMORTGGLFGSRRLVLTGNGDTPSGRSGSSEGGYGLNIEV
 ITGSGKSWKSEAEAGSGSSDTPGKXVYIIPPPEDDSI PAHYQYINPDYDTIV
 EVSGHDAFPAITFEEDNLCTANNIARKGKTKLPVQKRSIPIILARDIMACOT
 GSGRTAFLFLILAHMWDGILASRFELDPSCIIIVAPRELWNIYLAKEFSFGT
 CVRAVVIYGGTQLHMSIRQIVQCNIIICATPGRLMDIIGEKIQLQIYLVLEDR
 MLDWGFEPMKLLISCPGMSKEORQILMSAFPEIIOLEAFKSYLVLEVAQV
 GACRSDVQQTIVQGFSGKREKVEILRNIGDERTWVFEKKADPFAFPCQKES
 TTSIHGDECEEREQALGDFPGKCPVLVATSVYARGLDIENQVHINIDLSITDEY
 VHRIGTRCGNTRALSFDLSEDNHIAQPLVYKLTDAQODVPALWEIARSTYIPG
 FSGTRGVFPAVSDTRKXSLTNTAGSSSRANPVDDESDM"

ORIGIN

Query Match 100.0%; Score 2224; DB 9; Length 2224;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACTTGAAGTCAACATGGGGGATGAAGATGGGAAGCAGAAATACACCTCATATGCTTC 60
 Db 1 ACTTGAAGTCAACATGGGGGATGAAGATGGGAAGCAGAAATACACCTCATATGCTTC 60
 61 CTATGTCCTCATATTTTGAAGAGATAGATATTTCTGAGAAAATGGAACAATTTTAACG 120
 Db 61 CTATGTCCTCATATTTTGAAGAGATAGATATTTCTGAGAAAATGGAACAATTTTAACG 120
 121 GACTCAGCTCATCATATAGAAATGAGATGACCTCTCGAAGATCATTTCTGAA 180
 Db 121 GACTCAGCTCATCATATAGAAATGAGATGACCTCTCGAAGATCATTTCTGAA 180
 181 AAGTGATTTGCTCTGGGCGAAATTTTGAAGACAGAGATGCTGTGAGTGAATAAGCG 240
 Db 181 AAGTGATTTGCTCTGGGCGAAATTTTGAAGACAGAGATGCTGTGAGTGAATAAGCG 240
 241 AAGTGATTTGCTCTGGGCGAAATTTTGAAGACAGAGATGCTGTGAGTGAATAAGCG 240
 Db 241 AAGTGATTTGCTCTGGGCGAAATTTTGAAGACAGAGATGCTGTGAGTGAATAAGCG 240
 241 AGATTAATACATCCCAATGGGTGTTTGAAGTGAAGAGATTTTGAAGACAGAGATTT 300
 Db 241 AGATTAATACATCCCAATGGGTGTTTGAAGTGAAGAGATTTTGAAGACAGAGATTT 300
 301 TTCAAAACAGAGATTTGAAGATGATAGCTCTGCTTTCTGAGAGAGCTAGTAATGA 360
 Db 301 TTCAAAACAGAGATTTGAAGATGATAGCTCTGCTTTCTGAGAGAGCTAGTAATGA 360
 361 CTGCGAAGATTAATCCCAACAGAGATTTTCCCAAGAGAGGCGCTATCGAGATGG 420
 Db 361 CTGCGAAGATTAATCCCAACAGAGATTTTCCCAAGAGAGGCGCTATCGAGATGG 420
 421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAAAGAGTAGTTCCGAGGTG 480

421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAAAGAGTAGTTCCGAGGTG 480
 Db 421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAAAGAGTAGTTCCGAGGTG 480
 481 CCGTGAAGATTTGGTGTAGAAAGTCCAAATTAATGACTTGAAGCCAGAGAAATGATGA 540
 Db 481 CCGTGAAGATTTGGTGTAGAAAGTCCAAATTAATGACTTGAAGCCAGAGAAATGATGA 540
 541 GCGCACTGTGGCTTTTGTCTTGAAGACCAATTAATGATGTCACAGATTAATGATGA 600
 Db 541 GCGCACTGTGGCTTTTGTCTTGAAGACCAATTAATGATGTCACAGATTAATGATGA 600
 601 TACTTCCAAAGCAGAAAGTGGAGTGAACAGAGGTTTCAAAAGTTTAAATGA 660
 Db 601 TACTTCCAAAGCAGAAAGTGGAGTGAACAGAGGTTTCAAAAGTTTAAATGA 660
 661 AGAAGTAAATACAGGCTCTGAAAAGATTTCTTGAAGTCAAGAGCAAGAGAGAAAG 720
 Db 661 AGAAGTAAATACAGGCTCTGAAAAGATTTCTTGAAGTCAAGAGCAAGAGAGAAAG 720
 721 TAGTATACCTCAAGAGCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATGAGGA 780
 Db 721 TAGTATACCTCAAGAGCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATGAGGA 780
 781 CTCCATCTTTCACATTAATCAGACAGGCAATTAATCGAATAATACACATATTTCTGT 840
 Db 781 CTCCATCTTTCACATTAATCAGACAGGCAATTAATCGAATAATACACATATTTCTGT 840
 841 GGAAGTGTCTGGAATGATGACACACAGCAATTCGATTTTGAAGAGCTAATCTCTG 900
 Db 841 GGAAGTGTCTGGAATGATGACACACAGCAATTCGATTTTGAAGAGCTAATCTCTG 900
 901 TCAGACCTGAATTAACACATTTGCTAAGCTGTTTACTAAGCTTACTCTGTGCAAAA 960
 Db 901 TCAGACCTGAATTAACACATTTGCTAAGCTGTTTACTAAGCTTACTCTGTGCAAAA 960
 961 ATACAGATTCCTATCATATATTCAGACAGAGATTTGATGGCTTGCTCAACAGAGGTC 1020
 Db 961 ATACAGATTCCTATCATATATTCAGACAGAGATTTGATGGCTTGCTCAACAGAGGTC 1020
 1021 TGGGAAAGCTGCGGCTTTCTCTCAATTTTGGCTATATGATGATGATGAATTAAC 1080
 Db 1021 TGGGAAAGCTGCGGCTTTCTCTCAATTTTGGCTATATGATGATGATGAATTAAC 1080
 1081 TGGCAGTGTGTTTAAAGAGTGCAGAAACAGAGTATTAATGAGCCAACTCGAGA 1140
 Db 1081 TGGCAGTGTGTTTAAAGAGTGCAGAAACAGAGTATTAATGAGCCAACTCGAGA 1140
 1141 ATTGTCACCAAGATTTATTTGGAAGCCAGAAATTTCTTTGGAGCTTGTGAAGGC 1200
 Db 1141 ATTGTCACCAAGATTTATTTGGAAGCCAGAAATTTCTTTGGAGCTTGTGAAGGC 1200
 1201 TGTGTTATATATGAGGGAACCCAGCTGGACATTCATTTGGAACATTAATGATGATG 1260
 Db 1201 TGTGTTATATATGAGGGAACCCAGCTGGACATTCATTTGGAACATTAATGATGATG 1260
 1261 TAAATATATATGAGTACTCTGGAAGACTGATGATGATGATGATGATGATGATGATG 1320
 Db 1261 TAAATATATATGAGTACTCTGGAAGACTGATGATGATGATGATGATGATGATGATG 1320
 1321 TCTCAAAACAGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1380
 Db 1321 TCTCAAAACAGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1380
 1381 TGTTCAGAAATGAGAGATTAATTTCTGCGAGAAATGCGATCAAAAGAAACAGCGCA 1440
 Db 1381 TGTTCAGAAATGAGAGATTAATTTCTGCGAGAAATGCGATCAAAAGAAACAGCGCA 1440
 1441 AACCTTAATGTCAGTGAACCTTTCCAGAGAAATTCAAAGGTGCTGCTGAGATTTT 1500
 Db 1441 AACCTTAATGTCAGTGAACCTTTCCAGAGAAATTCAAAGGTGCTGCTGAGATTTT 1500
 1501 AAAGTCAATTAATCTGTTTCTGCTGTTGAGCAAGTGGTGAAGATGATGATGATGATG 1560

Db 1501 AAAAGTAAATATATCTGTTGTTGCTGTTGAGCAAGTGGGTGAGACATGTAGATGTTCA 1560
 QY 1561 GCAGACCGTTCTTCCAAAGTTGGCCAGTTCTCAAAAAGAGAAAAGCTCGTGAATTTGCGC 1620
 Db 1561 GCAGACCGTTCTTCCAAAGTTGGCCAGTTCTCAAAAAGAGAAAAGCTCGTGAATTTGCGC 1620
 QY 1621 AAACATAGGGGATGAAAGAACTATGCTGTTGTGAAAATAGAAAAGAGAGATTTTAC 1680
 Db 1621 AAACATAGGGGATGAAAGAACTATGCTGTTGTGAAAATAGAAAAGAGAGATTTTAC 1680
 QY 1681 TGGCACTTTCTTTGTCAAGAAAATATCAACTCAAGATATCAATGATGATCGGGAACA 1740
 Db 1681 TGGCACTTTCTTTGTCAAGAAAATATCAACTCAAGATATCAATGATGATCGGGAACA 1740
 QY 1741 GAGAGAGCGGAGCAAGCTCTTGAAGATTTTGGCTTTGAAAAGGCCAGTTCTGTTGC 1800
 Db 1741 GAGAGAGCGGAGCAAGCTCTTGAAGATTTTGGCTTTGAAAAGGCCAGTTCTGTTGC 1800
 QY 1801 TACTTCACTGCTGCCAGAGGCTGGATATGAAAATGTGCAACATGTTATCAATTTTGA 1860
 Db 1801 TACTTCACTGCTGCCAGAGGCTGGATATGAAAATGTGCAACATGTTATCAATTTTGA 1860
 QY 1861 TCTTCTTCTACCATTTGATGATATGTTTCATGAAATGGGCTAGTGGTGGTGGGAA 1920
 Db 1861 TCTTCTTCTACCATTTGATGATATGTTTCATGAAATGGGCTAGTGGTGGTGGGAA 1920
 QY 1921 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGAATCAATTTTGAACAGCTCT 1980
 Db 1921 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGAATCAATTTTGAACAGCTCT 1980
 QY 1981 AGTAAAGATTTGACAGATGCTCAACAGATGTTCTGATGATGATGATGATGATGATG 2040
 Db 1981 AGTAAAGATTTGACAGATGCTCAACAGATGTTCTGATGATGATGATGATGATGATG 2040
 QY 2041 TACTGATATCATTTCTGAGCTTCAAGTATGATGATGATGATGATGATGATGATGATG 2100
 Db 2041 TACTGATATCATTTCTGAGCTTCAAGTATGATGATGATGATGATGATGATGATGATG 2100
 QY 2101 TACCAAGAGGAGGAG 2160
 Db 2101 TACCAAGAGGAGGAG 2160
 QY 2161 TCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
 Db 2161 TCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
 QY 2221 TGGCA 2224
 Db 2221 TGGCA 2224

RESULT 2
 AF262962 2411 bp mRNA linear PRI 30-MAY-2000
 LOCUS Homo sapiens VASA protein mRNA, complete cds.
 DEFINITION AF262962
 ACCESSION AF262962.1 GI:8102020
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1. Rocha, D. and Affara, N.
 Cloning and characterization of the human VASA gene
 Unpublished
 2. (bases 1 to 2411)
 TITLE
 JOURNAL
 AUTHORS
 TITLE
 JOURNAL
 REFERENCES
 DIRECT SUBMISSION
 Submitted (02-MAY-2000) Department of Pathology, University of
 Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
 FEATURES
 source
 1. .2411

Query Match 99.6%; Score 2214.4; DB 9; Length 2411;
 Best Local Similarity 99.7%; Freq. No. 0;
 Matches 2218; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

ORIGIN

1 ACTTGAAGTCAACATGGGGATGAAAGATTGGAGAGCAAAATCAACCTCATATGCTTC 60
 5 ACTTGAAGCAACATGGAGATGAAAGATTGGAGAGCAAAATCAACCTCATATGCTTC 64
 61 CATATGTCATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
 65 CTATATGTCATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 124
 121 GACTCAGCTTCATCATGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 180
 125 GACTCAGCTTCATCATGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 184
 181 AGTGAATTTGCTTGGGCGGAATTTTGAAGAGAGATGCTGATGATGATGATGATGATG 240
 185 AGTGAATTTGCTTGGGCGGAATTTTGAAGAGAGATGCTGATGATGATGATGATGATG 244
 241 AGATATATCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 245 AGATATATCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 304
 301 TTCAAGAGAGAGATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 305 TTCAAGAGAGAGATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 364
 361 CTGCAAGATATCAACAG 420
 365 CTGCAAGATATCAACAG 424
 421 AAATATATCAAGAGATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 425 AAATATATCAAGAGATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
 481 CCGTGAAGATTTGCTTGAAG 540
 485 CCGTGAAGATTTGCTTGAAG 544
 541 GCGCACTGCTGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 545 GCGCACTGCTGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 604
 601 TACTTCTCAAG 660
 605 TACTTCTCAAG 664
 661 AGAAGATATACAGGCTCTGAAAGAGATTTCTGAAAGTCAAGAGAGAGAGAGAGAGAGAG 720

CDs

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 18..2192
 /codon_start=1
 /product="VASA protein"
 /protein_id="AF27205.1"
 /db_xref="GI:8102021"
 /translation="MDEDMBAEINPMSSVPIFEKDRYSGENDNPNTPASSEY
 DDGPERDRHEKSGFPASGRNFGNDADCECMKADJTSVWGFVGFSGFNGNSRFE
 DGSSEGFWRSESDCEDNPTNRGFSRGGYRGDNNSEASGSGSBRGSGFKRCRGP
 GIGSPNDLDPDECMORTGGLFGSRREPLVSGTGGDTSQSGSGSBRGSGKJLNBEV
 ITGSKNSWKSDEAGGESSDQGPRTVIPPPEDEDSIFAYQGTINPDKLITLV
 EVSGHADPAPLIPFEANLCOITLNNNAKAGYKLPVCKYSIPILAGRLMACQT
 GSKTAAFLPLIHAMHDDITARSRELEQPEECIIYAPTRVINOIYELARKRSPGT
 CVRAVPIYGTOLGHSIRQIVCGCNITCAPRMLDITIGEKTKGLKQIKVILDEADR
 MDMGFEPMKSLISCRGSPSKRQRLIMSATTPPEIOLABFLKSNLIPVAGQV
 GGACRDVQIVLVQVGFSSKRLKVEILRNIGDERTWVETKKAQPLATFLQEKIS
 TTSIHGREQREERQALGDRFGKCPVLVAARGLDIENQVHINFLPSTIDBY
 VHRIGTRRCGNTGRALSPFLSDNHLAQLVXVLTLDAQQDVPAMWLEIAFSTYIPG
 FSGTRGVFASVDTRKGSNTLNATGSSSQAPPVDESDM"

Db 665 AAGAGTATTAAGAGCTCTGAGAAAGATTCTTGAAAGTCAGAAACCAAGAGAGAAAG 724
 Qy 721 TAGTGATCTCAGAGACCAAAAGTACCTACATACCCCTCTCTCACTGAGATGAGGA 780
 Db 725 TAGATATCTCAGAGACCAAAAGTACCTACATACCCCTCTCTCACTGAGATGAGGA 784
 Qy 781 CTCACATCTTGGACATTAATCAGACAGGATTAACCTGACAAATACACACTATCTTGT 840
 Db 785 CTCACATCTTGGACATTAATCAGACAGGATTAACCTGACAAATACACACTATCTTGT 844
 Qy 841 GGAAGTGTGAGACATTAATCAGACAGGATTAACCTGACAAATACACACTATCTTGT 900
 Db 845 GGAAGTGTGAGACATTAATCAGACAGGATTAACCTGACAAATACACACTATCTTGT 904
 Qy 901 TCAGACATGATTAACCAATCTGTAAGCTGTATTAAGCTTACTCTCTGAGAAA 960
 Db 905 TCAGACATGATTAACCAATCTGTAAGCTGTATTAAGCTTACTCTCTGAGAAA 964
 Qy 961 ATACAGATTTCTATCAATCTTGCAGACAGATTTGATGCTTGTCTCAACAGGCTC 1020
 Db 965 ATACAGATTTCTATCAATCTTGCAGACAGATTTGATGCTTGTCTCAACAGGCTC 1024
 Qy 1021 TGGGAGAGCTGGGCTTTCTCTCAACATTTGGCTCAATGATGATGAGAAATAC 1080
 Db 1025 TGGGAGAGCTGGGCTTTCTCTCAACATTTGGCTCAATGATGATGAGAAATAC 1084
 Qy 1081 TGGGAGAGCTGGGCTTTCTCTCAACATTTGGCTCAATGATGATGAGAAATAC 1140
 Db 1085 TGGGAGAGCTGGGCTTTCTCTCAACATTTGGCTCAATGATGATGAGAAATAC 1144
 Qy 1141 ATTGATCAACAGATTTATTTGGAAGCAGAAAATTTCTTTGGGACTGTGTAAAGAC 1200
 Db 1145 ATTGATCAACAGATTTATTTGGAAGCAGAAAATTTCTTTGGGACTGTGTAAAGAC 1204
 Qy 1201 TGTGTATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1260
 Db 1205 TGTGTATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1264
 Qy 1261 TAAATATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1320
 Db 1265 TAAATATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1324
 Qy 1321 TGTCAACAGATCAAAATCTTAGTTTGGATGAGAAATTCGATGATGATGATGATG 1380
 Db 1325 TGTCAACAGATCAAAATCTTAGTTTGGATGAGAAATTCGATGATGATGATGATG 1384
 Qy 1381 TGTCCAGAAATGAAGAAGTTAATTTCTTCCAGAGATGCAATCAAGGAAACAGGCCA 1440
 Db 1385 TGTCCAGAAATGAAGAAGTTAATTTCTTCCAGAGATGCAATCAAGGAAACAGGCCA 1444
 Qy 1441 AACCTTATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1500
 Db 1445 AACCTTATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1504
 Qy 1501 AAAGCAAAATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1560
 Db 1505 AAAGCAAAATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1564
 Qy 1561 GAGAGACCTTCTCCAGATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1620
 Db 1565 GAGAGACCTTCTCCAGATGAGGAGAACCCAGCTGGGACATTCATTCGACAAATACCAAGCTG 1624
 Qy 1621 AAACATAGGAGAGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1680
 Db 1625 AAACATAGGAGAGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1684
 Qy 1681 TGCACATTTCTTGTCTCAAGAAAAATATCAACTCAAGATATCAAGATATCAAGATAT 1740
 Db 1685 TGCACATTTCTTGTCTCAAGAAAAATATCAACTCAAGATATCAAGATATCAAGATAT 1744
 Qy 1741 GAGAGAGGAGAGAGCTCTGAGAAATTTCTGTTGAAAGTCCCAAGTTCTTGTTC 1800

Db 1745 GAGAGAGGAGAGAGAGCTCTGAGATTTTGGCTTGGAAAGTGGCCAGTTCTTGTTC 1804
 Qy 1801 TACTTCTGATGCTCCAGAGAGGCTGATATGAAAATGCGAACATGATATGATTTGA 1860
 Db 1805 TACTTCTGATGCTCCAGAGAGGCTGATATGAAAATGCGAACATGATATGATTTGA 1864
 Qy 1861 TCTTCTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
 Db 1865 TCTTCTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1924
 Qy 1921 TACTGAGAGAGAGATTTCTTTTGTATCTGATGATGATGATGATGATGATGATGATG 1980
 Db 1925 TACTGAGAGAGAGATTTCTTTTGTATCTGATGATGATGATGATGATGATGATGATG 1984
 Qy 1981 AGTAAAGATATGACAGATGCTCAAGAGATGCTTCTGATGATGATGATGATGATGATG 2040
 Db 1985 AGTAAAGATATGACAGATGCTCAAGAGATGCTTCTGATGATGATGATGATGATGATG 2044
 Qy 2041 TAGTACATATATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 2100
 Db 2045 TAGTACATATATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 2104
 Qy 2101 TACCAGAAAGGAGAGAGAGCTTTGAAACAGAGCTGGGTTTCTTCTTCAAGAGCTCCCA 2160
 Db 2105 TACCAGAAAGGAGAGAGAGCTTTGAAACAGAGCTGGGTTTCTTCTTCAAGAGCTCCCA 2164
 Qy 2161 TCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
 Db 2165 TCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2224
 Qy 2221 TGCA 2224
 Db 2225 TGCA 2228

RESULT 2
 BC047455
 LOCUS
 DEFINITION
 Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 4, mRNA (cDNA clone MGC:48304 IMAGE:529807), complete cds.
 ACCESSION
 BC047455
 VERSION
 BC047455.1 GI:28958130
 KEYWORDS
 MGC.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, D., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshylyk, S., Caminci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunatene, P.H., Richards, S., Morken, K.C., Hale, S., Garcia, A.M., Gay, L.J., Holys, K.S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fehy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Schachenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skaleja, U., Smalins, D.E., Schercher, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 12477932
 REFERENCE
 2 (bases 1 to 2363)
 AUTHORS
 Strausberg, R.

TITLE

Direct Submission
Submitted (28-FEB-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunsatze, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

FEATURES

source

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ULNL at: <http://image.llnl.gov>
Series: IRAX Plates: 91 Row: k Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9507236.

Location/Qualifiers

1..2363
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:48304 IMAGE:529807"
/tissue_type="Testis"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
1..2363
/gene="DDX4"
/note="synonym: VASA"
/db_xref="LOCUSID:54514"
/db_xref="MIM:605281"
38..2110
/codon_start=1
/product="DDX4 protein"
/protein_id="AAH47455.1"
/db_xref="GI:28958131"
/db_xref="LOCUSID:54514"
/translation="MGEDWEABINPHMSVYPIFEKORYSGENDFNRTPASSSEM
DDGSRDHFMKSGFASGRNFGNDAGCNKRDNTSMGGGPGVGSKRNKGFNSRFE
DGSAGFWRESNDCENPTNRNGRPSKRDNDLPDCMCRGTGLPSRPPVYSGN
GDTSSRSRSGSSEGRGKGLADEVITGSGKSKMSKSEKSSSTGCPKTYIPPPP
EDSDIFAHVGTGINFEDKDTLIVESGHDAPPAITLFEENLQDTNNNTAAGATK
ITPVKYSIPILLAGRDLMACQOTGSGTAFLPILAHMMHGBITASRKELQEBEC
IIAPTRBLVANKIYLEARKFSRGTGCVRAVVIYGGTQGHISRQIVOCNLTCPSTL
MDIIKERKIGAKQIKYIVLEADRMIDMGFGEEMKLLISCPMSKSCROTLCATPST
PEIOLRLAEKFKSNVLFVAVGVGACRVOQTVLVGVGSKREKLEIVLITGER
TMVPEYTKKADFIATPLCOEKISITTSIHGDRERSEGLGDFRFGKCVLIVATVA
ARGLIENVGVHINFDLPSTIDVYHRIGRGRCNGNRAISPDLESNDHIAQPLVY
VITDNOQVPMLEBIARSTIIPGSSGSTRGNVAVSDTRKSKSTLNTAIPSSQAPN
PVDDSSWD
716..1915
/note="Smb; Region: Superfamily II DNA and RNA helicases
[DNA replication, recombination, and repair /
transcription / Translation, ribosomal structure and
biogenesis]"
/db_xref="CCD:CC050513"

CDS

Gene

misc_feature

ORIGIN

Query Match 89.8%; Score 1997.2; DB 9; Length 2363;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2114; Conservative 0; Mismatches 8; Indels 102; Gaps 1;

1 ACTTGAAGTCACACATGGGGATGAAGATTGGGAGCAGAAATCAACCTCATATGCTTC 60

25 ACTTGAAGTCACATGGGAGATGAAGATTGGGAGCAGAAATCAACCTCATATGCTTC 84
61 CTATGTTCCCATATTGGAAGATAGATATCTCGGAGAAATGGAACAAATTTAAACG 120
85 CTATGTTCCCATATTGGAAGATAGATATCTCGGAGAAATGGAACAAATTTAAACG 144
121 GACTCCAGCTTCATCATCAGAAATGATGATGATGATGATGATGATGATGATGATGAT 160
145 GACTCCAGCTTCATCATCAGAAATGATGATGATGATGATGATGATGATGATGATGAT 204
181 AAGTGGATTTGCTTGGGCGGAAATTTGGAAAGAGATGCTGGTGAAGTGAATGATG 240
205 AAGTGGATTTGCTTGGGCGGAAATTTGGAAAGAGATGCTGGTGAAGTGAATGATG 264
241 AGATATATCATTCACATGAGTGGTGGTTTGGAGTTGGAAGAGTTTGGAAACAGAGTT 300
265 AGATATATCATTCACATGAGTGGTGGTTTGGAGTTGGAAGAGTTTGGAAACAGAGTT 324
301 TTCAACAGCAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 360
325 TTCAACAGCAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 384
361 CTGCAAGATATATCAACAGCAGGATGATGATGATGATGATGATGATGATGATGATG 420
385 CTGCAAGATATATCAACAGCAGGATGATGATGATGATGATGATGATGATGATGATG 431
421 AATATATTCAGAGCTTCAGGCGCATACAGAGAGTGAAGATGATGATGATGATGATG 480
432 ----- 431
481 CCGTAGAGATTTGGTCTAGAGAGTCCAAATATGATGATGATGATGATGATGATGATG 540
432 -----ATAAGACTTGAACCCAGCAATGATGATGATGATGATGATGATGATGATG 462
541 GCGCATGCTGCTGCTTTTGGTTTGAAGACAGATTTTGAAGTGAAGTGAAGTGAAGTGA 600
463 GCGCATGCTGCTGCTTTTGGTTTGAAGACAGATTTTGAAGTGAAGTGAAGTGAAGTGA 522
601 TACTCTCAAGACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660
523 TACTCTCAAGACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 582
661 AGAATATATCAAGGCTTGAAGGATTTCTGGAAGTGAAGGATGAAGGATGAAGGATGA 720
583 AGAATATATCAAGGCTTGAAGGATTTCTGGAAGTGAAGGATGAAGGATGAAGGATGA 642
721 TAGTATCTCAAGGATGAAGGATTTCTGGAAGTGAAGGATGAAGGATGAAGGATGAAG 780
643 TAGTATCTCAAGGATGAAGGATTTCTGGAAGTGAAGGATGAAGGATGAAGGATGAAG 702
781 CTCATCTTTTGAAGATTTCAAGAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 840
703 CTCATCTTTTGAAGATTTCAAGAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 762
841 GGAATGCTGGAAGATTTCAAGAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGAAG 900
763 GGAATGCTGGAAGATTTCAAGAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGAAG 822
901 TGAAGACTGAATTAACAAGTGTGAAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 960
823 TGAAGACTGAATTAACAAGTGTGAAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 882
961 ATACAGTATCTCAATCAATCTGGAAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 1020
883 ATACAGTATCTCAATCAATCTGGAAGGATTTCTGGAAGTGAAGGATTTCTGGAAGTGA 942
1021 TGGGAAGCTGGGCTTTCTCTTCAACAATTTGGCTCATATGATGATGATGATGATGAT 1080
943 TGGGAAGCTGGGCTTTCTCTTCAACAATTTGGCTCATATGATGATGATGATGATGAT 1002
1081 TGGCAAGCTGTTTAAAGATTTGAGAGAACAGAGTGAATATGATGATGATGATGATGAT 1140

Db 1003 TCCAGTGTCTTTAAAGCTTGCAGAACCAAGAGTATATTGTACACCAACTGGA 1062
 Qy 1141 ATTGTCACACGATTTATTTTGAAGCCAGAAAAATTTCTTTGGGACTTGTGTAAGAC 1200
 Db 1063 ATGTGTCACAAAGATTTATTTTGAAGCCAGAAAAATTTCTTTGGGACTTGTGTAAGAC 1122
 Qy 1201 TGTGTGTATATATGGGGGAAACCCAGCTGGGACATTCATTCGCAATTAAGTATACAGGCTG 1260
 Db 1123 TGTGTGTATATATGGGGGAAACCCAGCTGGGACATTCATTCGCAATTAAGTATACAGGCTG 1182
 Qy 1261 TATATATATATATGTGTCTACTCTCTGGAAGAAGTATGATATATAGGCAAGAAAAAGATTG 1320
 Db 1183 TATATATATATATGTGTCTACTCTCTGGAAGAAGTATGATATATAGGCAAGAAAAAGATTG 1242
 Qy 1321 TCTCAACAGATTCGAATCTAGTTTGGATGAGTATGATGATGATGATGATGATGATGATGATG 1380
 Db 1243 TCTCAACAGATTCGAATCTAGTTTGGATGAGTATGATGATGATGATGATGATGATGATGATG 1302
 Qy 1381 TGTTCAGAAATGAAGAGATTATTTCTTCCAGAAATGATGATGATGATGATGATGATGATGATG 1440
 Db 1303 TGTTCAGAAATGAAGAGATTATTTCTTCCAGAAATGATGATGATGATGATGATGATGATGATG 1362
 Qy 1441 AACCTTATATGTCAGTGAACCTTTCCAGAGGAATTCAGAGTTGCTGACAGTTT 1500
 Db 1363 AACCTTATATGTCAGTGAACCTTTCCAGAGGAATTCAGAGTTGCTGACAGTTT 1422
 Qy 1501 AAAGTCAATTAATCTGTGTTGCTGTGGAAGAGGCTGAGATGATGATGATGATGATGATGATGATG 1560
 Db 1423 AAAGTCAATTAATCTGTGTTGCTGTGGAAGAGGCTGAGATGATGATGATGATGATGATGATGATG 1482
 Qy 1561 GGAGACGCTGCTCCAGTGGCCAGTGTCTCTCAAAAGAGAAAGAGTGTGTAATTTGGG 1620
 Db 1483 GGAGACGCTGCTCCAGTGGCCAGTGTCTCTCAAAAGAGAGAGTGTGTAATTTGGG 1542
 Qy 1621 AAACATAGGGAGTGAAGAACTATGCTTGTGTTGAACCTAAGAAAAAGCAGATTTTAC 1680
 Db 1543 AAACATAGGGAGTGAAGAACTATGCTTGTGTTGAACCTAAGAAAAAGCAGATTTTAT 1602
 Qy 1681 TGCACCTTTCTTGTCTCAAGAAAAATATCAACATACAGATGATGATGATGATGATGATGATGATG 1740
 Db 1603 TGCACCTTTCTTGTCTCAAGAAAAATATCAACATACAGATGATGATGATGATGATGATGATGATG 1662
 Qy 1741 GAGAGAGCGGAGAGAGCTCTTGAGAGATTTGCTGTTGGAAGTGGCCAGTTCTTGTG 1800
 Db 1663 GAGAGAGCGGAGAGAGCTCTTGAGAGATTTGCTGTTGGAAGTGGCCAGTTCTTGTG 1722
 Qy 1801 TACTTCAGTACGTCGAGAGGCTGATGATTTGAAGAAATGTCACATGATGATGATGATGATGATGATG 1860
 Db 1723 TACTTCAGTACGTCGAGAGGCTGATGATTTGAAGAAATGTCACATGATGATGATGATGATGATGATG 1782
 Qy 1861 TCTTCTCTTACCAATGATGATATGTTCAATGATGATGATGATGATGATGATGATGATGATG 1920
 Db 1783 TCTTCTCTTACCAATGATGATATGTTCAATGATGATGATGATGATGATGATGATGATGATG 1842
 Qy 1921 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGATTAACATTTTGAACAGCTCT 1980
 Db 1843 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGATTAACATTTTGAACAGCTCT 1902
 Qy 1961 AGTAAAGATATGACAGATGCTCAACAGATGTTCTCTGATGGTTGGAAGAAATTTGCTT 2040
 Db 1903 AGTAAAGATATGACAGATGCTCAACAGATGTTCTCTGATGGTTGGAAGAAATTTGCTT 1962
 Qy 2041 TAGATCAATCAATCTGCTGCTCAGTGTGATGATGATGATGATGATGATGATGATGATGATG 2100
 Db 1963 TAGATCAATCAATCTGCTGCTCAGTGTGATGATGATGATGATGATGATGATGATGATGATG 2022
 Qy 2101 TACAGAAAGGCAAGAGCACTTTGAAACAGAGTGGTGTCTTCTTCAAGAGCTGCCAA 2160
 Db 2023 TACAGAAAGGCAAGAGCACTTTGAAACAGAGTGGTGTCTTCTTCAAGAGCTGCCAA 2082
 Qy 2161 TCCAGTATATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
 Db 2083 TCCAGTATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2142

Qy 2221 TGCA 2224
 Db 2143 TGCA 2146
 RESULT 4
 HSM802178
 LOCUS HSM802178 2159 bp mRNA linear PRI 18-FEB-2000
 DEFINITION Homo sapiens mRNA; cDNA DKFZp434B1122 (from clone DKFZp434B1122);
 partial cds.
 ACCESSION AL137462
 VERSION AL137462.1 GI:6808043
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2189)
 AUTHORS Blum, H., Bauerhach, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 TITLE Direct Submission
 JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferplatz 18a, D-82152
 Martinsried, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Ludwig Maximilians University,
 Munich/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 This clone (DKFZp434B1122) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
 Location/Qualifiers
 1..2189
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp434B1122"
 /tissue_type="testis"
 /clone_id="434 (synonym: htess). Vector pSport1; host
 DH10B; sites="NotI + SalI"
 /dev_stage="adult"
 1..2189
 /gene="DKFZp434B1122"
 <1..1910
 /gene="DKFZp434B1122"
 /note="strong similarity to testis specific Helicase (VASA
 homolog)"
 /codon_start=3
 /product="hypothetical protein"
 /protein_id="CAB70750.1"
 /db_xref="GI:6808044"
 /db_xref="GOA:Q9NOI0"
 /db_xref="SWISS-PROT:Q9NOI0"
 /translation="SRGNRGSNSRPFDDSGFRRSSNDCENPTNRGFRGGY
 RDNNSRPSGPRYRGRGSGFRRCRGFGIGSNNDLPDRCMORPGI.FGRSRYLISG
 TNGDTSQSGSSGSGSRGKGLNREIVTGSKRWKSEABESSDYOGRKYTIIP
 PPEDSDSI.FAHYQTGINDKYDTLLIVEGSDAPALITFEANLCTLNNTAKAG
 YTKLTPVQKXSIPIILAGRDLMACAQSGSGKVAALPIILAHMEDIGTASRFELOE
 PECILVAPRELWQYILEARFSFGTCVRAVLYGTQLGHSISQIVOGSNILCATP
 GRMLDIIKREKIGLKOIKYLVDSEADRMIDMGFGEMKKLISCPMPSEKORQILMF
 ATPPEIOLAEFEKSNLFPVAVGVGAGCAQVQATLVQVGFSEKREKLEILNIG
 DERTWVFEVTKKKADRIATFELGQEKISITSSJHGDSERRERBOAGDPFGCPYLVA
 SVAAKRLDLENQHYINFDLPSTIDBYHRIERTRCNTRGALISFPLSDNHLAOP
 LVKVLTDQADVDVPAWLEETAFSTYIPGSGSIRGNVPAVSDTRKSKSLTNTAGRSSQ
 APNPVDSQSD"
 2128..2133
 /gene="DKFZp434B1122"
 2147
 /gene="DKFZp434B1122"
 polyA_signal
 polyA_site
 ORIGIN

Wed Jul 28 11:42:30 2004

us-09-714-865b-1.1.rge

Page 7

Query Match		87.2%:	Score 1939.6:	DB 9:	Length 2189:
Beet Local Similarity		99.8%:	Pred. No. 0:		
Matches 1942:		Conservative 0:	Mismatches 4:	Indels 0:	Gaps 0:
QY	279	AGAGTTTGGAAAAGAGTTTTCACACAGAGTTTGAAGATGATAGCTCTGTT			
DB	1	AGAGTTTGGAAAAGAGTTTTCACACAGAGTTTGAAGATGATAGCTCTGTT			
QY	339	TCTGAGAGATCTAGTATGATCTGCGAAGATATCCACACGGAACAGAGGTTTCCA			
DB	61	TCTGAGAGATCTAGTATGATCTGCGAAGATATCCACACGGAACAGAGGTTTCCA			
QY	399	AGAGAGCGGCTATCGAGATGAAATATTCAGAGGCTTCAGGCGCATACAGAGAGTG			
DB	121	AGAGAGCGGCTATCGAGATGAAATATTCAGAGGCTTCAGGCGCATACAGAGAGTG			
QY	459	GAAAGGATGTTTCCGAGGTTGCGTGGAGATTTGGTTCAGAGATCCCAATATGACT			
DB	181	GAAAGGATGTTTCCGAGGTTGCGTGGAGATTTGGTTCAGAGATCCCAATATGACT			
QY	519	TAGACCCAGACGAATGTATGACGCGCATGCTGCGCTTTTGGTTTCAAGACCAAT			
DB	241	TAGACCCAGACGAATGTATGACGCGCATGCTGCGCTTTTGGTTTCAAGACCAAT			
QY	579	TAACTGGACAGATATGATGATCTCTCAAGACGAATGCGAGTGAAGTGAACGAG			
DB	301	TAACTGGACAGATATGATGATCTCTCAAGACGAATGCGAGTGAAGTGAACGAG			
QY	639	GTGCTTCAAAAGTTTAAATGAAGATTAATACAGGCTCTGGAAGAAATCTTGAAGT			
DB	361	GTGCTTCAAAAGTTTAAATGAAGATTAATACAGGCTCTGGAAGAAATCTTGAAGT			
QY	699	CAGAGACGAGAGAGAGAGAGATGTATCTCAAGACCAAAAGTGAATCTTACATCCC			
DB	421	CAGAGACGAGAGAGAGAGAGATGTATCTCAAGACCAAAAGTGAATCTTACATCCC			
QY	759	CTCTCCACCTGAGATGAGATCTCATCTTGTGACATTAATCAGACGAGATTAATCTG			
DB	481	CTCTCCACCTGAGATGAGATCTCATCTTGTGACATTAATCAGACGAGATTAATCTG			
QY	819	ACAAATACGACATTAATCTTGTGAGATGTCTGACATGATGACCAACGACATTTCTA			
DB	541	ACAAATACGACATTAATCTTGTGAGATGTCTGACATGATGACCAACGACATTTCTA			
QY	879	CTTTTGAAGAGATTAATCTCTGACACATGATTAACACATTTGCTAAAGCTGTTATA			
DB	601	CTTTTGAAGAGATTAATCTCTGACACATGATTAACACATTTGCTAAAGCTGTTATA			
QY	939	CTAAGCTTACTCTCTGTCGAAAATACGATTTCTTATCATCTTGAAGAGAGATTTGA			
DB	661	CTAAGCTTACTCTCTGTCGAAAATACGATTTCTTATCATCTTGAAGAGAGATTTGA			
QY	999	TGGCTTGTGCTCAACAGGCTCTGGAAGATCTGCGCTTTCTCTACCAATTTGGCTC			
DB	721	TGGCTTGTGCTCAACAGGCTCTGGAAGATCTGCGCTTTCTCTACCAATTTGGCTC			
QY	1059	ATAATGATGATGATGATGATTAATCTGCGAGTCTTTTAAAGATTGACAGAGATGTA			
DB	781	ATAATGATGATGATGATGATTAATCTGCGAGTCTTTTAAAGATTGACAGAGATGTA			
QY	1119	TTATTTGAGACCAACTCGAGATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT			
DB	841	TTATTTGAGACCAACTCGAGATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT			
QY	1179	CTTTTGGAGCTTGTGTAAAGCTGTCTTATATATGAGGAGAACCCAGCTGGACATTTCA			
DB	901	CTTTTGGAGCTTGTGTAAAGCTGTCTTATATATGAGGAGAACCCAGCTGGACATTTCA			
QY	1239	TTTGCACAAATATGATCAAGGCTGTATATATATATATGCTCTCTGGAAGATGATGATA			
DB	961	TTTGCACAAATATGATCAAGGCTGTATATATATATATGCTCTCTGGAAGATGATGATA			
QY	1299	TCATATGCGAAAGAAAGATTTGCTCAAAACAGATCAATATCTTAGTTTGGATGAAGCTG			

DB	1021	TCATATGCGAAAGAAAGATTTGCTCAAAACAGATCAATATCTTAGTTTGGATGAAGCTG			
QY	1359	ATCCGATGTTGATATGATGATTTTGGTCAGAAATGAAGAGTTATTTCTTGCCAGGAA			
DB	1081	ATCCGATGTTGATATGATGATTTTGGTCAGAAATGAAGAGTTATTTCTTGCCAGGAA			
QY	1419	TGCATCAAAAGAAACAGGCGCAACCTTATGTTGCTGCAATTTTCCAGAGAAATTC			
DB	1141	TGCATCAAAAGAAACAGGCGCAACCTTATGTTGCTGCAATTTTCCAGAGAAATTC			
QY	1479	AAAGGTTGCTGCAAGATTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGATG			
DB	1201	AAAGGTTGCTGCAAGATTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGATG			
QY	1539	GTGAGATGATGAGATGTTTCAAGACCTTCTCAAGTTGSCCATTTCTCAAAAGAG			
DB	1261	GTGAGATGATGAGATGTTTCAAGACCTTCTCAAGTTGSCCATTTCTCAAAAGAG			
QY	1599	AAAGGCTGTTGAATTTCTGCAAGATGAGGAGTAAAGACTATGCTTTTGTGAAA			
DB	1321	AAAGGCTGTTGAATTTCTGCAAGATGAGGAGTAAAGACTATGCTTTTGTGAAA			
QY	1659	CTAAGAAAAAGCAGATTTTACTGCAATTTCTTTGTCAAGAAAAATATCACTACAA			
DB	1381	CTAAGAAAAAGCAGATTTTACTGCAATTTCTTTGTCAAGAAAAATATCACTACAA			
QY	1719	GTATCATGTTGATTCGGGAAACAGAGAGCGGAGCAAGCTTTGAGATTTTCTGTTG			
DB	1441	GTATCATGTTGATTCGGGAAACAGAGAGCGGAGCAAGCTTTGAGATTTTCTGTTG			
QY	1779	GAAAGTCCCAAGTTCTTTGCTTACTTCAATGCTGCGAGAGGCTGCAATTTGAAAAG			
DB	1501	GAAAGTCCCAAGTTCTTTGCTTACTTCAATGCTGCGAGAGGCTGCAATTTGAAAAG			
QY	1839	TGCAACATGTTATCAATTTTATGATCTTCTTCAATGATGATGATATGTTCAATGAA			
DB	1561	TGCAACATGTTATCAATTTTATGATCTTCTTCAATGATGATGATATGTTCAATGAA			
QY	1889	GGCGTATGCTGCTGTTGGGAAATATGCGACAGCAATTTCTTTTGTATCTTGAATCG			
DB	1621	GGCGTATGCTGCTGTTGGGAAATATGCGACAGCAATTTCTTTTGTATCTTGAATCG			
QY	1959	ATAACATTTAGACAGGCTCTAGTAAAGATTTGACAGATGCTCAACAGATGTTCTG			
DB	1681	ATAACATTTAGACAGGCTCTAGTAAAGATTTGACAGATGCTCAACAGATGTTCTG			
QY	2019	CATGTTGGAAGAAATTTGCTTTATGATATCAATTTCTGCTTCAAGTGTATGACAG			
DB	1741	CATGTTGGAAGAAATTTGCTTTATGATATCAATTTCTGCTTCAAGTGTATGACAG			
QY	2079	GAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
DB	1801	GAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
QY	2139	TTTCTTCTTCAAGATCTCCATTCAGATGATGATGATGATGATGATGATGATGATG			
DB	1861	TTTCTTCTTCAAGATCTCCATTCAGATGATGATGATGATGATGATGATGATGATG			
QY	2199	TCTTCAAGTCTGTTGATGCA 2224			
DB	1921	TCTTCAAGTCTGTTGATGCA 1946			

RESULT 5
AX748127 1984 bp mRNA 1linear PART 20-JUN-2003
LOCUS AX748127
DEFINITION Sequence 1652 from Patent EP1308459.
ACCESSION AX748127
VERSION AX748127.1 GI:32132515
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuko, Y.

TITLE

Full-length cDNA sequences
Patent: EP 1308459-A 1652 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)

FEATURES

Location/Qualifiers

1..1984

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 72.0%; Score 1601; DB 6; Length 1984;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1662; Conservative 0; Mismatches 5; Indels 48; Gaps 1;

QY 510 ATATGACTAGACCCAGAGAAATGTAATGAGGCACTGGTGGCTTTTGGTTCTAGAA 569
DB 117 ATATGACTAGACCCAGAGAAATGTAATGAGGCACTGGTGGCTTTTGGTTCTAGAA 176
QY 570 GACCAATATTAGTGGCAAGTATGATGATCTTCTCAAGCAAGAGTGGCACTGCA 629
DB 177 GACCAATATTAGTGGCAAGTATGATGATCTTCTCAAGCAAGAGTGGCACTGCA 236
QY 630 GTGAAGAGGTGGTTCAAAAGTTTAAATGAAGAAATTAACAGGCTTGAAAGAAAT 689
DB 237 GTGAAGAGG-----ATT 248
QY 690 CTGGAAGTCAAGAGAGAGAGAGAAAGTATGATCTCAAGCAAGCAAGCAAGTACT 749
DB 249 CTGGAAGTCAAGAGAGAGAGAGAGAAAGTATGATCTCAAGCAAGCAAGTACT 308
QY 750 AACTACCCCTCTCTCACTGAGATGAGACTCCATCTTTCACATTAATGACAGCA 809
DB 309 AACTACCCCTCTCTCACTGAGATGAGACTCCATCTTTCACATTAATGACAGCA 368
QY 810 TAACTTCGCAAAATGACACATATCTTGGAAGTCTGGAATGATGACACACAG 869
DB 369 TAACTTCGCAAAATGACACATATCTTGGAAGTCTGGAATGATGACACACAG 428
QY 870 CAATTCGACTTTTGAAGAGCTAAATCTCTGACAGACATGAATTAACAATTTGCTAAG 929
DB 429 CAATTCGACTTTTGAAGAGCTAAATCTCTGACAGACATGAATTAACAATTTGCTAAG 488
QY 930 CTGTTATATCTAAGCTTCTCTGCAAAAATTAAGATTCCTATCTACTGACAGAC 989
DB 489 CTGTTATATCTAAGCTTCTCTGCAAAAATTAAGATTCCTATCTACTGACAGAC 548
QY 990 GAGATTGATGGCTTGCTCAACAGAGTCTGGAGAGACAGGCTTTTCTCTACCA 1049
DB 549 GAGATTGATGGCTTGCTCAACAGAGTCTGGAGAGACAGGCTTTTCTCTACCA 608
QY 1050 TTTTGGCTCATATGATGATGATGATTAATCTGCGAGTCTTTTAAAGTGTGAGAAC 1109
DB 609 TTTTGGCTCATATGATGATGATGATTAATCTGCGAGTCTTTTAAAGTGTGAGAAC 668
QY 1110 CAGAGTATATATGATGACCAACTCGAGAAATGATCAACCAATTAATTTTGAAGCA 1169
DB 669 CAGAGTATATATGATGACCAACTCGAGAAATGATCAACCAATTAATTTTGAAGCA 728
QY 1170 GAAATTTTCTTTGGAGCTTGTAAGAGCTGTTTATATATGAGGAGAAACCCAGCTGG 1229
DB 729 GAAATTTTCTTTGGAGCTTGTAAGAGCTGTTTATATATGAGGAGAAACCCAGCTGG 788
QY 1230 GACATTCATTCGACAAATAGTACAGAGCTGATATATATATGAGTACCTCGAGAGAC 1289
DB 789 GACATTCATTCGACAAATAGTACAGAGCTGATATATATATGAGTACCTCGAGAGAC 848

QY 1230 TGATGATATCATAGCCAAAGAAAGATTTGCTCAACAGATCAATTAATTAATTTGG 1349
DB 849 TGATGATATCATAGCCAAAGAAAGATTTGCTCAACAGATCAATTAATTAATTTGG 908
QY 1350 ATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1409
DB 909 ATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968
QY 1410 GCCCAGAAATCCATCAAG 1469
DB 969 GCCCAGAAATCCATCAAG 1028
QY 1470 AGGAATTCAGAAAGTGGCTGCAAGATTTTAAAGTCAAAATTAATCTGTTGGTCTGG 1529
DB 1029 AGGAATTCAGAAAGTGGCTGCAAGATTTTAAAGTCAAAATTAATCTGTTGGTCTGG 1088
QY 1530 GACAGTGGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589
DB 1089 GACAGTGGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148
QY 1590 CAAAAGAGAAAGCTGTTGAATTTCTGCAACATGAGGAGTGAAGAGATGATGAT 1649
DB 1149 CAAAAGAGAAAGCTGTTGAATTTCTGCAACATGAGGAGTGAAGAGATGATGAT 1208
QY 1650 TTGTTGAATCTAAGAAAGAGAGATTTTCTGCACTTTCTTGTCAAGAAATAT 1709
DB 1209 TTGTTGAATCTAAGAAAGAGAGATTTTCTGCACTTTCTTGTCAAGAAATAT 1268
QY 1710 CAATCAAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
DB 1269 CAATCAAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1328
QY 1770 TTGGCTTTGGAAGTGGCCAGTCTTGTCTCTCACTGATGATGATGATGATGATGAT 1829
DB 1329 TTGGCTTTGGAAGTGGCCAGTCTTGTCTCTCACTGATGATGATGATGATGATGAT 1388
QY 1830 TTGMAATGTCACAAATGATCAATTTGATCTTCTCTCACTGATGATGATGATGAT 1889
DB 1389 TTGMAATGTCACAAATGATCAATTTGATCTTCTCTCACTGATGATGATGATGAT 1448
QY 1890 ATGGAATTTGGGCTTACTGCTGCTTGGGAATCTGCGAGAGCAATTTCTTTTGTATC 1949
DB 1449 ATGGAATTTGGGCTTACTGCTGCTTGGGAATCTGCGAGAGCAATTTCTTTTGTATC 1508
QY 1950 TTGAATCGATTAACCAATTTAGCAGAGCTCTGATTAAGTATGACAGATGCTCAACAG 2009
DB 1509 TTGAATCGATTAACCAATTTAGCAGAGCTCTGATTAAGTATGACAGATGCTCAACAG 1568
QY 2010 ATGTTCTGATGCTTGAAGAAATGCTTTAGTATACATCAATTCCTGCTTCACTGAT 2069
DB 1569 ATGTTCTGATGCTTGAAGAAATGCTTTAGTATACATCAATTCCTGCTTCACTGAT 1628
QY 2070 GTACAGAGAGAAACGTTTGGATCAGTATGATGATGATGATGATGATGATGATGAT 2129
DB 1629 GTACAGAGAGAAACGTTTGGATCAGTATGATGATGATGATGATGATGATGATGAT 1688
QY 2130 CAGCTGGTCTTCTTCTGACAGAGCTCCCAATCAGATGATGATGATGATGATGAT 2189
DB 1689 CAGCTGGTCTTCTTCTGACAGAGCTCCCAATCAGATGATGATGATGATGATGAT 1748
QY 2190 GCCAABACATCTCTCAAGTCTGCTGTTTGAATCA 2224
DB 1749 GCCAABACATCTCTCAAGTCTGCTGTTTGAATCA 1783

RESULT 6
AK093439 1984 bp mRNA linear PRI 15-JUL-2002
LOCUS Homo sapiens cDNA FLJ36120 f1s, clone T85T12022738, highly similar
DEFINITION
CO DEAD BOX PROTEIN 4.
ACCESSION
AK093439
VERSION AK093439.1 GI:21752310

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
oligo capping; fis (full insert sequence)	Homo sapiens (human)			oligo capping; fis (full insert sequence)	Homo sapiens (human)		
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1	Nishi,T., Oda,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Isono,Y., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Irie,R., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fuji,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isegai,T.	NEDO human cDNA sequencing project		Unpublished			
2 (bases 1 to 1984)	Isegai,T. and Yamamoto,J.	Direct Submission		Submitted (04-JUL-2002) Takao Isegai, FUJ Project (HRI Team) ; 2-6-7			
	Kaatsa-Kanetari, Kizatsura, Chiba 292-0812, Japan (E-mail:genom@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)						
	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) / cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.						
location/Qualifiers							
1..1984							
/organism="Homo sapiens"							
/mol_type="mRNA"							
/db_xref="taxon:9606"							
/clone="TEST12022738"							
/tissue_type="testis"							
/clone_lib="TEST12"							
/note="Cloning vector: pME18F13"							
ORIGIN							
Query Match	72.0%;	Score 1601;	DB 9;	Length 1984;			
Best Local Similarity	96.9%;	Pred. No. 0;					
Matches 1662;	Conservative	0;	Mismatches	5;	Indels	48;	Gaps 1
510 ATATGACCTTAAGACCCAGACGAAATGATGACGACGACCTGGGCTTTGGTCTAGAA							569
117 ATATGACCTTAAGACCCAGACGAAATGATGACGACGACCTGGGCTTTGGTCTAGAA							176
570 GACCACTTAAATGAGTGCACAGGTAAATGATGATCTCTCAAGCAGAAAGTGCAGTGGAA							629
177 GACCACTTAAATGAGTGCACAGGTAAATGATGATCTCTCAAGCAGAAAGTGCAGTGGAA							236
630 GTGACGAGGTGTTCAAAAGTTTAAATGAGAAAGTAATACAGGCTCTGGAAGAAT							689
237 GTGACGAGG-----ATT							248
690 CTGGAAGTGCAGAAAGCAGAAAGTGTGATCTCAAGACCAAAAGTGACCT							749
249 CTGGAAGTGCAGAAAGCAGAAAGTGTGATCTCAAGACCAAAAGTGACCT							308
750 ACATACCCCTCTCTCACTGAGATGAGACCTCCATCTTTGACATTTACAGACGGCA							809
309 ACATACCCCTCTCTCACTGAGATGAGACCTCCATCTTTGACATTTACAGACGGCA							368
810 TAAATCTGCAATATGACACATATCTTGTGAAAGTGTGACATGATGACACCAACAG							869
369 TAAATCTGCAATATGACACATATCTTGTGAAAGTGTGACATGATGACACCAACAG							428
870 CAATTGCACTTTGAAAGAGCTAATCTGTGACAGACCTGAATTAACACATTGCTTAAG							929
429 CAATTGCACTTTGAAAGAGCTAATCTGTGACAGACCTGAATTAACACATTGCTTAAG							488
930 CTGGTTACTAGCTTCTGTGCAAAATACATATTCCATCATCTTGACAGGAC							989

[illegible]

ORIGIN

Qy	253	ACCTGAAGCTATCATGGGAGATGAAGATTGGAAAGACGAAT---	CAACCCCTCATATGTC	5
Qy		ACTTGAAGTCCACATGGGGSATGAAGATTGGAAAGACGAAT---	CAACCCCTCATATGTC	5
Db		TTCTCATGTTCCCAATATTTGAAAGATAGTGA--TTCTGGA	GAATAATGAGACAATTT	11
Qy	58	TTCTCATGTTCCCAATATTTGAAAGATAGTGA--TTCTGGA	GAATAATGAGACAATTT	11
Db		TTCTCATGTTCCCTGTATTTGAGAAGATATAATTTCTTCTG	GGGCAAAATGAGACAATTT	37
Qy	313	TTCTCATGTTCCCTGTATTTGAGAAGATATAATTTCTTCTG	GGGCAAAATGAGACAATTT	37
Qy	115	TAACGAGCTCCAGCTTCATCATCAGAAATGAGATGAGACTTC	TCGACAGATCATTT	17
Db		TAACGAGCTCCAGCTTCATCATCAGAAATGAGATGAGACTTC	TCGACAGATCATTT	17
Qy	373	TAACGAGCTCCAGCTTCATCATCAGAAATGAGATGAGACTTC	TCGACAGATCATTT	43
Qy	175	CATGAAAAGTGGATTGGCTCTGGGGGGAATTTGGAAA	CAGAGATGCTGTGATGTGA	23
Db		CATGAAAAGTGGATTGGCTCTGGGGGGAATTTGGAAA	CAGAGATGCTGTGATGTGA	23
Qy	433	CATGAAAAGTGGATTGGCTCTGGGGGGAATTTGGAAA	CAGAGATGCTGTGATGTGA	49
Qy	235	TAACGAGATPATATCAATCCACAATGGGTGGTTTGGAGTGA	AGTAAGTAATTTGGAAAACAG	29
Db		TAACGAGATPATATCAATCCACAATGGGTGGTTTGGAGTGA	AGTAAGTAATTTGGAAAACAG	29
Qy	493	TAAAGAGAGACTACATCTACACCGGTGGCTTTGGAAGAGAA	GGGATTTTGGAAAACAG	55
Db		TAAAGAGAGACTACATCTACACCGGTGGCTTTGGAAGAGAA	GGGATTTTGGAAAACAG	55
Qy	295	AGGTTTTTCAACAGCAGGTTTGAAGATGATAGCTGTGGTCT	GTGAGAGAGCTAG	35
Db		AGGTTTTTCAACAGCAGGTTTGAAGATGATAGCTGTGGTCT	GTGAGAGAGCTAG	35
Qy	553	AGGTTTTTAAATPACAAGTTTGAAGAGTGAACAGCTCTGG	TTTTCGAAAAGTCTAC	61
Db		AGGTTTTTAAATPACAAGTTTGAAGAGTGAACAGCTCTGG	TTTTCGAAAAGTCTAC	61
Qy	355	TAATGATCGGAGATPATATCCACAACCGGAACGAGAGTTC	TCCAAAGAGAGCGGCTATCG	41
Db		TAATGATCGGAGATPATATCCACAACCGGAACGAGAGTTC	TCCAAAGAGAGCGGCTATCG	41
Qy	613	TAATGATCGGAGATPATATCCACAACCGGAACGAGAGTTC	TCCAAAGAGAGCGGCTATCG	67
Db		TAATGATCGGAGATPATATCCACAACCGGAACGAGAGTTC	TCCAAAGAGAGCGGCTATCG	67
Qy	415	AGATGAAATPATATTCAGAGCTTCAGGGCATACAGAGAGTGA	AGAGGTATTTCCG	47
Db		AGATGAAATPATATTCAGAGCTTCAGGGCATACAGAGAGTGA	AGAGGTATTTCCG	47
Qy	673	AGATGGAATGATTCGGAAGCTTCAGGCCATTCGAAAGAGTGG	AGAG-----	72
Db		AGATGGAATGATTCGGAAGCTTCAGGCCATTCGAAAGAGTGG	AGAG-----	72
Qy	475	AGGTGCCCTGGAGGATTTGGTCTAGGAAGTCCAAATATAGT	ACTTAAACCCACAGATG	53
Db		AGGTGCCCTGGAGGATTTGGTCTAGGAAGTCCAAATATAGT	ACTTAAACCCACAGATG	53
Qy	723	-----ATAGTAATATATACAAAGATCAAGG	747	
Db		-----ATAGTAATATATACAAAGATCAAGG	747	
Qy	535	TATGACGCGCATGCTGGGCTTTTGGTGTCTAGAAGACAGTAT	TAAATGTCACAGATA	59
Db		TATGACGCGCATGCTGGGCTTTTGGTGTCTAGAAGACAGTAT	TAAATGTCACAGATA	59
Qy	748	ATCACAGCGGTGGTGGCTTTTGGTCTTAGGAACACAGACAA	GTGATTCAGGACG	80
Db		ATCACAGCGGTGGTGGCTTTTGGTCTTAGGAACACAGACAA	GTGATTCAGGACG	80
Qy	595	TGGTGAATCTTCTCAAGCAGAGTGGCAGTGAAGTGAACGA	GTGATTTCAAAGGTTT	65
Db		TGGTGAATCTTCTCAAGCAGAGTGGCAGTGAAGTGAACGA	GTGATTTCAAAGGTTT	65
Qy	808	TGGTGAATCTTCTCAAGCAGAGTGGCAGTGAAGTGAACGA	GTGATTTCAAAGGTTT	86
Db		TGGTGAATCTTCTCAAGCAGAGTGGCAGTGAAGTGAACGA	GTGATTTCAAAGGTTT	86
Qy	655	AAATGAAGAAGTATATACAGGCTCTGGAAGAAATTTCTTGA	AGTCAAGACAGAGAGAGG	71
Db		AAATGAAGAAGTATATACAGGCTCTGGAAGAAATTTCTTGA	AGTCAAGACAGAGAGAGG	71
Qy	862	AAATGAAGAAGTATATACAGGCTCTGGAAGAAATTTCTTGA	AGTCAAGACAGAGAGAGG	92
Db		AAATGAAGAAGTATATACAGGCTCTGGAAGAAATTTCTTGA	AGTCAAGACAGAGAGAGG	92
Qy	715	AGAAATAGTATCTCAAGACCAAAAGTACCTAATACCCCTCTC	CACTTGAGGA	74
Db		AGAAATAGTATCTCAAGACCAAAAGTACCTAATACCCCTCTC	CACTTGAGGA	74
Qy	922	CGAAACAGATGATATCAAGGTCCAAAGTACATATATACCC	CTCTCCACACAGGGA	98
Db		CGAAACAGATGATATCAAGGTCCAAAGTACATATATACCC	CTCTCCACACAGGGA	98
Qy	775	TGAGAGCTCCATCTTTTGGACATTAATCAGACAGGATATA	ACTTCGACAAATATGACATAT	83
Db		TGAGAGCTCCATCTTTTGGACATTAATCAGACAGGATATA	ACTTCGACAAATATGACATAT	83
Qy	982	TGAGAGCTCCATCTTTTGGACATTAATCAGACAGGATATA	ACTTCGATTAATATGATACAT	104
Db		TGAGAGCTCCATCTTTTGGACATTAATCAGACAGGATATA	ACTTCGATTAATATGATACAT	104
Qy	835	TCCTTGGAAAGTCTGACATGATGACCAACAGCAATCTGAT	CTTTTGAAGAACTAA	89
Db		TCCTTGGAAAGTCTGACATGATGACCAACAGCAATCTGAT	CTTTTGAAGAACTAA	89
Qy	1042	ACTTGTGAAGTATCTGACATGATGACCAACCGGCAATTT	TGACTTTTGAAGAAAGGA	110
Db		ACTTGTGAAGTATCTGACATGATGACCAACCGGCAATTT	TGACTTTTGAAGAAAGGA	110
Qy	895	TCTCTGTCAAGACACTGATTAACAACATTTGCTTAAAGCT	GTATTACTAGCTTACTCTGT	95
Db		TCTCTGTCAAGACACTGATTAACAACATTTGCTTAAAGCT	GTATTACTAGCTTACTCTGT	95
Qy	1102	TCTCTGTCAAGACACTGATTAACAACATTTGCTTAAAGCT	GTATTACTAGCTTACTCTGT	116
Db		TCTCTGTCAAGACACTGATTAACAACATTTGCTTAAAGCT	GTATTACTAGCTTACTCTGT	116
Qy	955	GCAAAAATACATATCTCTATCATACTTGCAGAGCGAGATT	ATGATGGCTGTGCTCAAC	101
Db		GCAAAAATACATATCTCTATCATACTTGCAGAGCGAGATT	ATGATGGCTGTGCTCAAC	101
Qy	1162	GCAGAAAGTACAGCAATCTCCATGTGTAGCAGAAAGATTT	GATATGGCTGT	

Db 1282 AATATACGACAGTGGCTTTAAAGAACTGACAGAAACCAAGATGATATATATGATACCAAC 1341
 Qy 1135 TCGAAGATTGGTCAACCAAGATTTATTTGGAAGCCAGAAAATTTCTTTGGGACTGGT 1194
 Db 1342 TCGAAGATTGATCAACCAAAATTTATTTGGAAGCCAGAAAATTTCTTTGGGACTGGT 1401
 Qy 1195 AAGAGCTGTGTATATATATGAGGAAACCCAGCTGGGACATTCATTGACAAATATGACA 1254
 Db 1402 AAGAGCTGTGTATATATGAGGAAACCCAGCTGGGACATTCATTGACAAATATGACA 1461
 Qy 1255 AAGCTGTATATATATATATGAGGAAACCCAGCTGGGACATTCATTGACAAATATGACA 1314
 Db 1462 AAGCTGTATATATATATATGAGGAAACCCAGCTGGGACATTCATTGACAAATATGACA 1521
 Qy 1315 GATTGGTCTGAAACAGATCAAACTACTAGTTTGGATGAGAGTGTGATGATAT 1374
 Db 1522 GATTGGTCTGAAACAGATCAAACTACTAGTTTGGATGAGAGTGTGATGATAT 1581
 Qy 1375 GGGTTTGTGTCAGAAATGAGAAAGTTAATTTCTTGGCCAGAAATGCCATCAAGAAACA 1434
 Db 1582 GGGTTTGTGACCCGAAATGAGAAAGTTAATTTCTTGGCCAGAAATGCCATCAAGAAACA 1641
 Qy 1435 GCGCCAAACCTTATGTTCACTGCACTTTCCAGAGAAATTCAAAAGTTGGCTGAGA 1494
 Db 1642 GCGCCAAACCTTATGTTCACTGCACTTTCCAGAGAAATTCAAAAGTTGGCTGAGA 1701
 Qy 1495 GTTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGAGTGGTGAAGATGAGAGA 1554
 Db 1702 GTTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGAGTGGTGAAGATGAGAGA 1761
 Qy 1555 TGTTCAGAGACCGTTCTCCAGTTGGCCAGTTCTCAAAAAGAGAAAAGCTCGTTGAAT 1614
 Db 1762 TGTTCAGAGACCGTTCTCCAGTTGGCCAGTTCTCAAAAAGAGAAAAGCTCGTTGAAT 1821
 Qy 1515 TCTGCGAAACATAGAGGAGTGAAGAACTATGCTTTGTTGAAATTAAGAAAAGAGAGA 1674
 Db 1822 TCTGCGAAACATAGAGGAGTGAAGAACTATGCTTTGTTGAAATTAAGAAAAGAGAGA 1881
 Qy 1675 TTTTACTGCAACTTTTCTTTGTCAAGAAAATATCAACTATCAAGTATTCATGCGAGTCG 1734
 Db 1882 TTTTACTGCAACTTTTCTTTGTCAAGAAAATATCAACTATCAAGTATTCATGCGAGTCG 1941
 Qy 1735 GGAACAGAGAGAGGGGAGCAAGCTTTGGAGATTTTGGCTTGGAAAGTGGCCAGTTCT 1794
 Db 1942 GGAACAGAGAGAGGGGAGCAAGCTTTGGAGATTTTGGCTTGGAAAGTGGCCAGTTCT 2001
 Qy 1795 TGTGCTACTTCAGTAGCTGCGACAGAGGCTTGATATGAAAATGTGCAACATGTTATCAA 1854
 Db 2002 TGTGCTACTTCAGTAGCTGCGACAGAGGCTTGATATGAAAATGTGCAACATGTTATCAA 2061
 Qy 1855 TTTTGATTTCTCTTCTACCATTTGATGAATATGTTCAATGAGTGGGAGTCTGCTG 1914
 Db 2062 TTTTGATTTCTCTTCTACCATTTGATGAATATGTTCAATGAGTGGGAGTCTGCTG 2121
 Qy 1915 TGGGAATCTGCGACAGAGCAATTTCTTTTGTGATCTTGAATCGAATAACATTTAGACA 1974
 Db 2122 TGGGAATCTGCGACAGAGCAATTTCTTTTGTGATCTTGAATCGAATAACATTTAGACA 2181
 Qy 1975 GCGCTTACTGAAAAGTATGACAGATGCTCAACAGAGATTTCTGATGATGTTGGAAGAAAT 2034
 Db 2182 ACGCTTACTGAAAAGTATGACAGATGCTCAACAGAGATTTCTGATGATGTTGGAAGAAAT 2241
 Qy 2035 TGGCTTTAGTACAT---CATTCCTGGCTCAGTGTGATCAAGAGGAAACGTGTTGC 2091
 Db 2242 TGGCTTTAGTACAT---CATTCCTGGCTCAGTGTGATCAAGAGGAAACGTGTTGC 2301
 Qy 2092 ATCAGTTGATACAGAA-----AAGGCAAGAGCACTTTGAACACAGCTGGGATTTTC 2142
 Db 2302 ATCTTTTGACACTAGAGAAATTTCCAGGGCAAGAAACACAGTGGGATTTTC 2361
 Qy 2143 TTCTTCAGAGCTCCCAATCCAGTAGATGATGATGAGTCAAGGATTTAAAGCA--TAAACATC 2200

Db 2362 TTTCGACAGCTCCCAATCCAGTTGATGACGAGTCAATGAGATTTAAAGCAACAGACATA 2421
 Qy 2201 CTTCAAGTCTGTGGTTT 2218
 Db 2422 CTGCAAGCTGTATGTTT 2439
 RESULT 8
 MUSDVH 1930 bp mRNA linear ROD 05-FEB-2003
 LOCUS Mus musculus mRNA for Drosophila vasa homologue, partial cds.
 ACCESSION D14859
 VERSION D14859.1 GI:286074
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 TITLE
 Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H.,
 Furusawa, M., and Noce, T.
 Isolation of a DEAD-family protein gene that encodes a murine
 homolog of Drosophila vasa and its specific expression in germ cell
 lineage
 Proc. Natl. Acad. Sci. U.S.A. 91 (25), 12258-12262 (1994)
 MEDLINE 95083681
 PUBMED 7991615
 REFERENCE
 PUBLISHED 2 (bases 1 to 1930)
 AUTHORS
 TITLE
 Direct Submission
 Submitted (01-APR-1993) Yoshiaki Noce, Mitsubishi Kasei Institute
 of Life Sciences, Developmental Biology, 11 Minamiooya, Machida,
 Tokyo 194, Japan (Tel:81-427-24-6246, Fax:81-427-29-1252)
 FEATURES
 source
 1..1930
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="BALB/c"
 /sub_species="domesticus"
 /db_xref="taxon:10090"
 /cell_type="primordial germ cell, spermatogonium,
 spermatocyte"
 /tissue_type="gonad, testis"
 /clone_lib="lambda gt10"
 /dev_stage="adult, embryo"
 <1..1914
 /function="RNA helicase"
 /codon_start=1
 /product="Drosophila vasa homologue"
 /protein_id="BAA03584.1"
 /db_xref="GI:286075"
 /translation="RGRGKGFGNRRGFLNNKFEKSPGSKWNNDCEDNDRSGFS
 KGGGCDNDNSASGPRRGGRGGRGGLRPNSESPODGTGCGGFFVLGX
 PNASDSNGDITQSRSGSGRGYKLNREVVTGSGKNSKSTEGESDSGGRVTV
 IPPPEDEDSDFAHYQTGINFDKDTLLIVESGDAPAILTFEBAICQTLNNIR
 KAGYTLTPVQKXTTPIVLADGLMAACAQGTGKDAFLPLILAHMMDGTASFKK
 LQEPICILVAPRELINQIYLBARKFSFCTCVISVYIGTGFSGSVRIYGCNLIK
 ATGRLMDIIGKXIKQVTLVLEADSDMDGFAPEIKLIKSCPGMSEKHQTL
 LSAAPREIQRADGFLKSNYLVAVGVGGACDPAVOQTILOVGVGKESKLLRFE
 NIGDEKTMFVETKKKADFIATFLCEKISSTISGDRPRRERDGLDFRGKCPVL
 VASVAAKLDLENVDVHNPDPSTIDETVHRIGRTGRCNTGRALISFDPSDNHL
 ADPLVVLSDAQQDVPAMLEBIAFTYVPSPSSSTRGSAVPAVSVDTRKNTGKARHE
 YSGDFEFTSSQSS"
 ORIGIN
 Query Match 64.9%; Score 1443; DB 10; Length 1930;
 Best Local Similarity 86.4%; Pred. No. 0;
 Matches 1671; Conservative 0; Mismatches 240; Indels 24; Gaps 6;
 Qy 266 TTTCAGTTGGAAGAAGTTTGGAAACAGAGCTTTTCAACAGCAGGTTTGAAGATGGT 325
 Db 1 TTTCGAAAGGAAAGGCTTTGGAACAGAGCTTTTAAATPAAAGTTTGAAGAAGGT 60

QY 326 GATAGCTCTGGTTCTGAGAGAGTCTGATGATATGATCGGAGATTAATCCAAACCGGAC 385
 Db 61 GATAGCTCTGGTTCTGAGAGAGTCTGATGATATGATCGGAGATTAATCCAAACCGGAC 120
 QY 386 AGAGGCTTTTCCAGAGAGCGGCTATCGAGATGGAATATTATAGAGGCTTCAGGCGCA 445
 Db 121 AGAGGCTTTTCCAGAGAGCGGCTATCGAGATGGAATATTATAGAGGCTTCAGGCGCG 180
 QY 446 TACGAGAGAGTGGAGAGAGTGTGTTCCGAGGTGCGGTGAGAGATTGGTCTAGGAAGT 505
 Db 181 TTCGAGAGAGGCGGAGAGGAGGAGTTCGAGGCTGCGGTGAGAGATTGGTCTAGGAAGT 240
 QY 506 CCAATATATGATGATGAGAGAGAGATGATGAGAGGAGTGTGCGGCTTTTGTGCT 565
 Db 241 CCAATATGATGATGATGAGAGAGAGATGATGAGAGGAGTGTGCGGCTTTTGTGCTA 300
 QY 566 AGAGAGCAGTATTAGTGGACAGTATGATGATATCTTCAAGAGAGAGTGGCAGT 625
 Db 301 GGGAGAGCAGCAGAGAGTGTGATGAGAGATGATGAGAGATGATGAGAGAGAGAGAGT 354
 QY 626 GGAAGTGAAGAGAGTGTGATGAGAGAGTGTGATGAGAGAGTGTGATGAGAGAGAGT 685
 Db 355 GGAAGTGTGAGAGTGTGATGAGAGAGTGTGATGAGAGAGTGTGATGAGAGAGAGT 414
 QY 686 AATTCCTGGAAGTGAAGAGAGAGAGAGAGAGAGTGTGATGATGAGAGAGAGAGAGT 745
 Db 415 AATTCCTGGAAGTGAAGAGAGAGAGAGAGAGAGTGTGATGATGAGAGAGAGAGAGT 474
 QY 746 ACCATCATACCCCTCTCCAGCTGAGAGATGAGAGATCCATCTTTGACATTAATGACAA 805
 Db 475 ACATATATACCCCTCTCCAGCTGAGAGATGAGAGATCCATCTTTGACATTAATGACAA 534
 QY 806 GGCATTAACCTGACAAATAGACACATATCTGTGGAAGTGTGATGAGAGAGAGAGT 865
 Db 535 GGCATTAACCTGATTAATATGACACATATCTGTGGAAGTGTGATGAGAGAGAGAGT 594
 QY 866 CAGCAATCTGACTTTTGAAGAGAGTATCTGTGACAGACTGATTAACCAATGCT 925
 Db 595 CCGGCAATTTGACTTTTGAAGAGAGTATCTGTGACAGACTGATTAACCAATGCT 654
 QY 926 AAGGCTGTATATCTAGCTTACCTGTCGCAAAATACAGTATTCCTTCAATCTGTA 985
 Db 655 AAGGCTGTATATCTAGCTTACCTGTCGCAAAATACAGTATTCCTTCAATCTGTA 714
 QY 986 GAGCAGATTTGATGAGGCTGTGCTCAAAAGGCTGTGGAAGAGAGTGTGCTCTA 1045
 Db 715 GAGCAGATTTGATGAGGCTGTGCTCAAAAGGCTGTGGAAGAGAGTGTGCTCTA 774
 QY 1046 CCAATTTGGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1105
 Db 775 CCAATTTGGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 834
 QY 1106 GAAACAGAGTGTATTTGATGAGACCACTCGAGAAATGATGATGATGATGATGAT 1165
 Db 835 GAAACAGAGTGTATTTGATGAGACCACTCGAGAAATGATGATGATGATGATGAT 894
 QY 1166 GCCAGAAATTTTCTTTGGAGCTTGTGAAGAGCTGTGATATATATATATATATAT 1225
 Db 895 GCCAGAAATTTTCTTTGGAGCTTGTGAAGAGCTGTGATATATATATATATATAT 954
 QY 1226 CTGGACATTTCAATTCGACAAATAGTACAGAGGCTGATATATATATATATATAT 1285
 Db 955 TTGTCATTCATGATTCGACAGATAGTACAGAGGCTGATATATATATATATATAT 1014
 QY 1286 AGACTGATGATATCATATGAGCAAAAGAAAGATTTGCTCAACAGATCAATATCTAGT 1345
 Db 1015 AGGCTATGATGATCATATGAGCAAAAGAAAGATTTGCTCAACAGATCAATATCTAGT 1074
 QY 1346 TTGATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1405
 Db 1075 TTGATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134

QY 1406 TCTTCCCGAGAGATTCATCAAGAGAGAGGCGCAACCTTATGTTGATGCAATTTT 1465
 Db 1135 TCTTCCCGAGAGATTCATCAAGAGAGAGGCGCAACCTTATGTTGATGCAATTTT 1194
 QY 1466 CCAGAGAGAAATCAAGAGTGTGCTGACAGTGTGTTTAAAGTAAATATATATATAT 1525
 Db 1195 CCAGAGAGAAATCAAGAGTGTGCTGACAGTGTGTTTAAAGTAAATATATATATAT 1254
 QY 1526 GTTGAACAAGTGTGAGCATATGAGATGTTGACAGAGAGGCTTCCAGATTTGGCCAG 1585
 Db 1255 GTTGGCAAGTGTGAGCATATGAGATGTTGACAGAGAGGCTTCCAGATTTGGCCAG 1314
 QY 1586 TTCTCAAAAAGAGAAAGCTGTTGAAATTTCTGGG-AAACATAGGAGATGAAACATAT 1644
 Db 1315 -TATCAAAAAGAGAAAGCTGTTGAAATTTCTAGGAAACATATAGTATGAAACATAT 1373
 QY 1645 GGTCTTTGTTGAAATCAAGAAAGAGAGATTTTACTGCAATTTTCTTTGTCAAGAAA 1704
 Db 1374 GGTCTTTGTTGAAATCAAGAAAGAGAGATTTTACTGCAATTTTCTTTGTCAAGAAA 1433
 QY 1705 AATATCACTACAGATATTCATGATGATGATGATGATGATGATGATGATGATGAT 1764
 Db 1434 AATATCAAGTACAGATATTCATGATGATGATGATGATGATGATGATGATGATGAT 1493
 QY 1765 AGATTTTGTGTTGAAAGTCCAGTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1824
 Db 1494 AGATTTTGTGTTGAAAGTCCAGTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1553
 QY 1825 GGAATTTGAAATGTGCAACATGTTATCAATTTGATCTTCTTCAACATGATGATA 1884
 Db 1554 TGATATGAAATGTGCAACATGTTATCAATTTGATCTTCTTCAACATGATGATA 1613
 QY 1885 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1944
 Db 1614 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1673
 QY 1945 TGAATCTGAATGAGATTAACATTTTGAACAGAGCTTCTAGTAAAGTATGACAGATGCTA 2004
 Db 1674 TGAATCTGAATGAGATTAACATTTTGAACAGAGCTTCTAGTAAAGTATGACAGATGCTA 1733
 QY 2005 ACAGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2061
 Db 1734 ACAGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1793
 QY 2062 CAGTGTAGTACAG 2109
 Db 1794 CAGTGTAGTACAG 1853
 QY 2110 -GGCAG 2168
 Db 1854 GGGCAG 1913
 QY 2169 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2183
 Db 1914 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1928

RESULT 9
 AB004836 1989 bp mRNA linear VRT 01-SEP-2000
 LOCUS
 DEFINITION Gallus gallus mRNA for Cvh, complete cds.
 ACCESSION AB004836
 VERSION AB004836.1 GI:9967267
 KEYWORDS
 SOURCE
 ORGANISM
 Gallus gallus (chicken)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (sites)
 Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T. and Noce, T.
 Isolation of chicken vasa homolog gene and tracing the origin of
 primordial germ cells

JOURNAL MEDLINE 20283536
 PUBMED 10821771
 2 (sites)
 Tsunekawa, N., Fujimoto, H., Nishida, T. and Noce, T.
 Isolation of a DEAD-family protein gene that encodes a chicken homolog of Drosophila vasa and its specific expression in germ cell lineage
 JOURNAL Unpublished
 3 (bases 1 to 1989)
 Tsunekawa, N.
 Direct Submission
 Submitted (13-JUN-1997) Naoki Tsunekawa, Mitsubishi Kasei Institute of Life Sciences, Molecular Reproduction and Development; 11 Minamiooya, Machida, Tokyo 194, Japan
 (E-mail: tsunek@libra.1s.m.kagaku.co.jp, Tel: +81-427-24-6246, Fax: +81-427-24-6316)
 Location/Qualifiers
 1. 1989
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 1. 1989
 /codon_start=1
 /product="Cvtn"
 /protein_id="BAB12337.1"
 /db_xref="GI:9667268"
 /translation="MEEDMDTELEQENAAAGSREDOAMNANGSRNPSLRFSSRP
 SPSLSPGPRNPSFPGFSONKSLANEINLSLPVQHDIGSGSRESVYRNED
 OPIYRGRGRSGSRDROENRNSANDPQMDQGRVYVGIQSGCNSENRNPLGS
 PPAFGKGVGVPAGVTKRSEIDSRGKTYVPPRPEDRQSI PACYOSINPX
 YDEAVMSGLDPPAPLAFEEANFAQLKMSIKTYGSKLTVQGHSLPVIDAGDL
 MSAQTSGLSGLAFLEPIVDRMKDVTAFPRQDPQCIIVATSELINQIFLEAK
 FVYGTICRPVIVGGTQTHSIRQIMQGCNIIATPQRLDIIEKRSISVEKYVL
 DEADRLMDGFGIDMKCLISYPMPSKDRQTLMPFATPEEYORLAGEFLKDYFL
 VIGNTGAGSDVQONILQVPRLSKRLIILQSTGGERMTVVDTRKKADYLAFIC
 QENLPSTSHGDRQRETRALSDPRSGKQILVATSVASRGIDINQVHINFDLPN
 TIEDYVRIGTRGTCNTGKAVSFDDQSGHLVQSLIKVLSLTQDFPQSGMAVQR
 TNIVASTWCPKGLQAVAEWNPFRMRKYSSETTFKWE"

ORIGIN
 Query Match 33.8%; Score 751.2; DB 5; Length 1989;
 Best Local Similarity 71.1%; Pred. No. 1,7e-167;
 Matches 1009; Conservative 0; Mismatches 408; Indels 3; Gaps 1;

Db 661 AGAAGTAATAACAGGCTCTGGAAAGAAATCTTGAAGTCAGAGCAGAGAGAGAGAAAG 720
 492 AG 551
 721 TAGGATATCTCAAG 780
 552 TGAATCTGGAAG 611
 781 CTCGATCTTTGCAATATACAG 840
 612 GTCCATCTTTGCAATATACAG 671
 841 GGAAGTCTGCAATATACAG 900
 672 TGAAGTCTGCAATATACAG 731
 901 TCAGACATCTGCAATATACAG 960
 732 TCAGACATCTGCAATATACAG 791
 961 ATACAGATCTGCAATATACAG 1020
 792 GCACAGCATCTCTGTTATACAG 851
 1021 TGGAG 1080
 852 AG 911

1081 TGCCAGTCGTTTAAAGAGTTGCAGAGACAGAGTGTATTTGTACACCAACTCGAGA 1140
 912 TGCAAGC---TTCCAAAGCAG 968
 1141 ATTGTCACAG 1200
 969 ACTGATTAATCAGATCTTCTTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1028
 1201 TGTGTTATATATAGGAG 1260
 1029 TGTGTTATATATAGGAG 1088
 1261 TAAT 1320
 1089 TAAT 1148
 1321 TCTCAAG 1380
 1149 TTTGTTGAG 1208
 1381 TGTGTCAG 1440
 1209 TGATTTAGATATAG 1268
 1441 AACCTTAT 1500
 1269 AACATTAAT 1328
 1501 AAAGTCAT 1560
 1329 GAAAG 1388
 1561 GAG 1620
 1389 GCAAAAT 1448
 1621 AAACATAG 1680
 1449 AGCAG 1508
 1681 TGCAACTTTTCTTTGTCAG 1740
 1509 TCCAG 1568
 1741 GAG 1800
 1569 GAG 1628
 1801 TACTTCAGTATGTCAG 1860
 1629 AACTTCAGTATGTCAG 1688
 1861 TCTTCTCTTACATATGAG 1920
 1689 TCTTCTCTTACATATGAG 1748
 1921 TACTTCAGTATGTCAG 1980
 1749 TACTTCAGTATGTCAG 1808
 1981 AGTAAAGATATGAG 2040
 1809 ACTTAAAGATATGAG 1868
 2041 TAGTACATATCTCTGCTGTCAG 2080
 1869 AGAACAATATATGTTGTTTCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1908

RESULT 10
 AF046043 2502 bp mRNA linear VEST 19-FEB-1998
 LOCUS AF046043
 DEFINITION Xenopus laevis DEAD box protein (VLG1) mRNA, complete cds.

Qy 1855 TTTTGATCTTCTCTTACCATTTGTAATGTCATGAAATGGCGGTAAGTGTGCTTG 1914
 Db 1879 TTTTGACATGCGCCAGCGACATGATGATGTCATCGATAGAGAACTGGAAGCTG 1938
 Qy 1915 TGGGAATATCTGGCAGAGCAATTCCTTTTGTGATCTTGAATCGGATTAACATTACCA 1974
 Db 1939 TGGGAACACCGGTGCGCGGTGCTTTTAAACCCAGAGTCTGACATCTCACTAGCTG 1998
 Qy 1975 GCTCTAGTAAAGTATTTGACAGATGCTCAACAGAGATGTTCTGCAATGTTGGAAGAAAT 2034
 Db 1999 CTCCTGTGCAAAATCTTTTCAGGGGCCCAAGAGTTCCAAAATGCGCTGAGAGAGT 2058
 Qy 2035 TGCTTAGTACATACATCTCTGCTCAGTGTGATAGACAAGAGAAAGCTGTTCCATC 2094
 Db 2059 TGCTTAGTGTCTTA---TGACACACAGAGCTTCAACCCAGTGAAGAGTGTTCATC 2115
 Qy 2095 AGTTGATACCAAGAAAGG 2112
 Db 2116 TACAGACTCGCGAAGG 2133

RESULT 12
 LOCUS DRY12007 2492 bp mRNA linear VRT 09-OCT-1997
 DEFINITION D. rerio vlg (vasa like gene) mRNA for putative RNA helicase.
 ACCESSION Y12007
 VERSION Y12007.1 GI:2558534
 KEYWORDS RNA helicase; vasa gene; vlg gene.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE
 AUTHORS Olsen, L.C., Aasland, R. and Fjose, A.
 TITLE A vasa-like gene in zebrafish identifies putative primordial germ cells
 JOURNAL Mech. Dev. 66 (1-2), 95-105 (1997)
 MEDLINE 98025484
 PUBMED 9376327
 REFERENCE Olsen, L.C.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-1997) L.C. Olsen, University of Bergen, Department of Molecular Biology, Thormoehensgt 55, N-5020 Bergen, NORWAY

FEATURES
 source location/Qualifiers

1..2492
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /dev_stage="embryonic, late blastula"
 1..2492
 /gene="vlg"
 35..2138
 /gene="vlg"
 /note="vasa-like gene
 putative"
 /codon_start=1
 /evidence=not experimental
 /product="RNA helicase (DEAD box)"
 /protein_id="CAA72735.1"
 /db_xref="GI:2558535"
 /db_xref="GOA:O42378"
 /db_xref="SPTREMBL:O42378"

/translation="MDWEDQSPVAVSGSGFGAGANDKSNSEGTGSSWNTGDSFR
 GRGGRGSGRGRGSGRGSFKSBIIDENSGDGNNGSRGRGRGFRGSRGSRDENE
 NNDNDGKSGRGRGGRGFRGGRGPDGDNEDTGRGRGRNENNGENDEGGEG
 RGRGRGRGGRGGRGGRGGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGR
 YRPPRPRRPSRPSHYATGNTFPRYDILIDVSGSNPKAINTFEAGLSDSKV
 SSGYKPTPVGKGLPIISGRDLMAQAQSGSKTAFLPLPIQRFRTDVAASKVS
 EIQEPRAIIAFTRELINQIYLEARKFAVGTCPVPPVYGGINTGYTIREVLKGCNVL

CATPGRUHLIDIGKXIGISKRYVLVLEADRMIDMGEPEMERKLVASPGMSKEKQ
 LMFATYPERDIQMAADFLKVDYFLVAVGVGACSDVEQITVQVDYSRDQLELL
 RATNERMTVETTERKSRADLAFPLCOEKISTHSIHQDREKREKALSPRLGHCY
 LVATSVARGDIDVOHVNFDMPSIDIEVNHIGTGRCNGTGRVSPFNPSDTP
 LARSLVYLSGAQOVPRKMLEEVAFSHGTTGFNPRKVASTDTRSGSKRDEPDP
 SOTSRPSSAAADBEWE"
 2255..2277
 /gene="vlg"
 /note="similar to nanos response element"
 /bound_moiety="a putative nanos-dependent factor"
 /evidence=not experimental

Query Match 30.9%; Score 586.4; DB 5; Length 2492;
 Best Local Similarity 69.6%; Pred. No. 4.2e-152;
 Matches 927; Conservative 0; Mismatches 405; Indels 0; Gaps 0;
 Db 715 AGAAGTATGATTAATCAAGACCAAAAGTACATATACCCCTCCTCACCTGAGGA 774
 656 AGAAGTATGATTAATCAAGACCAAAAGTATGATGATGATGATGATGATGATGAT 715
 Qy 775 TGAGACTCCATCTTTGCACTTATCAAGACGATAACTTGCAAAATACGACTAT 834
 Db 716 GGAAGCTCGATATTCCTCCATTATGCAACAGGCAATTTGACAAATATGATATAT 775
 Qy 835 TCTTGTGAAGTGTCTGACATGATGACACCAAGCAATTCGATCTTTGAAGAGCTAA 894
 Db 776 TCTTGTGAAGTGTCTGACATGATGACACCAAGCAATTCGATCTTTGAAGAGCTAA 835
 Qy 895 TCTGTGTGACACTGATTAACCAATGCTTAAGCTGTATATTAAGCTTACTCTGT 954
 Db 836 ACTTGTGACACTGATTAACCAATGCTTAAGCTGTATATTAAGCTTACTCTGT 895
 Qy 955 GCAAAATATGATTAATCAAGACCAAAAGTATGATGATGATGATGATGATGATGAT 1014
 Db 896 TCAAAATATGATTAATCAAGACCAAAAGTATGATGATGATGATGATGATGATGAT 955
 Qy 1015 AGGCTCTGGAAGACTGCGCTTTTCTCTACCAATTTGCTCATATGATGATGATGAT 1074
 Db 956 TGATCAAGAAAGAGCGCGCTTCTGCTGCCATCTTCAAGCGCTTATGATGATGAT 1015
 Qy 1075 AATACAGTCCAGTCTTTTAAAGTGTGCAAGACCAAGATGATATTTATGACACCA 1134
 Db 1016 TGATCAAGAAAGAGCGCGCTTCTGCTGCCATCTTCAAGCGCTTATGATGATGAT 1075
 Qy 1135 TCGAAGTGTGATCAAGATTAATTTGAAGCAGAAATTTCTTTGGAGCTGTGT 1194
 Db 1076 CAGAACTTATCAATGATCTATCTGAGGCCAGAAATTTGATGACGAGACCTGTGT 1135
 Qy 1195 AAGAGCTGTGATTAATATGAGGGAACCAAGCTGGAATTTCAATTCGACAAATGATCA 1254
 Db 1136 ACGACCTGTGATTAATATGAGGGAACCAAGCTGGAATTTCAATTCGACAAATGATCA 1195
 Qy 1255 AGGCTGTATATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1314
 Db 1196 GGGCTGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1255
 Qy 1315 GATTGCTGCAAAAGATCAATTAATTTGATGATGATGATGATGATGATGATGATGAT 1374
 Db 1256 GATTGCTGCAAAAGATCAATTAATTTGATGATGATGATGATGATGATGATGATGAT 1315
 Qy 1375 GGGTTTGTGCAAAAGATCAATTAATTTGATGATGATGATGATGATGATGATGATGAT 1434
 Db 1316 GGGTTTGTGCAAAAGATCAATTAATTTGATGATGATGATGATGATGATGATGATGAT 1375
 Qy 1435 GCGCAAAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494
 Db 1376 ACGCAAAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1435
 Qy 1495 GTTTTAAAGTCAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1554
 Db 1436 CTTTCTTAAGTCAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1495


```

Db 1218 TGTGTTATGAGAGTAAATACCGGATATCTATCCGAGAGGTGTTAAAGGCGTCAA 1277
Oy 1264 TATATTAATGCTCTCTCTGGAAGACTGATGATATATATAGGCAAAAGATTTGCTC 1323
Db 1278 TGTATGTCCTCTCTCTGGAAGACTGATGATATATAGGCAAAAGATTTGCTC 1337
Oy 1324 CAAACAGATCAATTAATTAATTTTGGATGAAAGCTGATGCGATG-----TTGATATGGG 1377
Db 1338 GAGTAAAGTGGCTATCTAGTTCTGATGAAAGCAGACAGATGCTGACATGACATGGG 1397
Oy 1378 TTTTGGTCCAGAAATGAAAGATTAATTTCTGCGCCAGAAATGCCATCAAGAGACAGCG 1437
Db 1398 CTGAGAGCCAGATATGCTAGTAAAGCTAGAGGATCTCCGGGATGCTTCAAAAGAGATCG 1457
Oy 1438 CCAACCCCTATGTTCAAGTCAATTTTCCAGAGAAATCAAAAGTTGCTGCAAGTT 1497
Db 1458 ACAACCCCTATGTTCAAGTCAATTTTCCAGAGAAATCAAAAGTTGCTGCAAGTT 1517
Oy 1498 TTTAAAGTCAATTAATTAATTTTGGTCTGTTGCAAAAGGCTGAGAGCATGAGATGT 1557
Db 1518 TTTAAAGTCAATTAATTAATTTTGGTCTGTTGCAAAAGGCTGAGAGCATGAGATGT 1577
Oy 1558 TCACAGACCGTCTCCAAAGTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTCCT 1617
Db 1578 GAGCAAACTATCGTTAGAGTAC--AGTACCAAGAGACAGCTGTTGAATTCCT 1634
Oy 1618 GCGAAATATGAGGAGTAAAGAACTATGTTGTTGTTGAAACTAGAGAAAGACAGATT 1677
Db 1635 CAGATCTACAGAGGATTAAGCAACAAATGCTTGTGTTGAAACCAAGAGAGTCTATTT 1694
Oy 1678 TACTGCAACTTTCTTTGTCAGAGAAATATCACTACAGATTCATGCTGATGCTGGGA 1737
Db 1695 CATTCGCAATTTCTCTGTCAGAGAGATGCTTACTACAGATTCATGCTGATGCTGGA 1754
Oy 1738 ACAAGAGAGCGGAGAGCAAGCTTTGAGATTTTCTGTTGAAAGTGGCCAGTTCTTGT 1797
Db 1755 GCAACGAGAGCGAGAGAGGCTCTCAATGATTTTCCGACAGGCTGCTGATGCTGCT 1814
Oy 1798 TGTCTACTGATGCTGCTGAGAGGCTGATATTTGAAATGTCGACATGTTATCAATT 1857
Db 1815 GGCAGAGTCTGCTGCTGAGAGGCTGATGATGAGAGTCCAGAGCTGCTGATGATTT 1874
Oy 1858 TGAATCTCTCTTACCATGATGATGATGATGATGATGATGATGATGATGATGATG 1917
Db 1875 CGAAGTCCAGAGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATG 1934
Oy 1918 GATTAAGTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1977
Db 1935 GAAATCTGAGAGGAGGCTGCTGTTTAAACCAAGAGTCTGACATCCATTAAGCTGCTC 1994
Oy 1978 TCTAGTAAAGTATGACAGATGCTCAACAGAGTCTGATGATGATGATGATGATGATG 2037
Db 1995 TCTGCTCAAGATGCTGCTGAGAGGCTCAACAGAGTCTGATGATGATGATGATGATG 2054
Oy 2038 CTTTAAATCAATTAATTAATTTTGGTCTGATGATGATGATGATGATGATGATGATG 2097
Db 2055 TTTCAATGCTCA--TGAAACCAAGATGATGATGATGATGATGATGATGATGATG 2111
Oy 2098 TGAATCAAGAGAGGAGAGAGCACTTTGAAACAGTGGGTTTCTTTCAAGAGCTTC 2157
Db 2112 AGACACCGAGAGAG--AGGCTCTTCAAGAGAGTGGCGCTCAAGAGCTGAGAGCTC 2168
Oy 2158 CAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2217
Db 2169 AAGCCAGAGAGGCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 2228
Oy 2218 T 2218
Db 2229 T 2229

```

RESULT 14

```

AB051835 2058 bp mRNA linear VRT 14-FEB-2002
LOCUS AB051835
DEFINITION Oreochromis niloticus vas-s mRNA for vasa short form, complete cds.
ACCESSION AB051835
VERSION AB051835.1 GI:14211583
KEYWORDS
SOURCE
ORGANISM
Oreochromis niloticus (Nile tilapia)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Neanthopterygii; Perciformes; Perciformes;
Labroidae; Cichlidae; Oreochromis.
REFERENCE
1 Kobayashi, T., Kajiyama-Kobayashi, H. and Nagahama, Y.
Two isoforms of vasa homologs in a teleost fish: their differential
expression during germ cell differentiation
Mech. Dev. 111 (1-2), 167-171 (2002)
JOURNAL
MEDLINE
11804791
PUBMED
2 (bases 1 to 2058)
REFERENCE
Kobayashi, T.
Direct Submission
Submitted (27-NOV-2000) Tohru Kobayashi, National Institute for
Basic Biology, Department of Developmental Biology, Myodaiji-cho,
Okazaki, Aichi 444-8585, Japan [E-mail: tohru_k@nib.ac.jp,
Tel:81-564-55-7554, Fax:81-564-55-7556]
FEATURES
source
1..2058
/organism="Oreochromis niloticus"
/mol_type="mRNA"
/db_xref="taxon:8128"
/clone="vas-s"
/clone_id="lamda zapII tilapia ovary cDNA"
1..2058
/gene="vas-s"
48..1913
/gene="vas-s"
/codon_start=1
/product="vasa short form"
/protein_id="BAB5110.1"
/db_xref="GI:14211584"
/translation="MDMEDEQFATSGTCALTSDTGTGCTGCPMNSDHGFGGGRGG
GGRGGGFGKNSPFGDQGNVTAERSFSFSGRABEDDKNVSGRRGRGGGG
GGRGGGFGGRGDEKSLKARGVPANLNASRPVTVVPPSPSDEDSIFSVSG
VNEPKDELIVDSGNTPPAINTPPEBALCELSKRVSSGTVKTPYOKSIEPILS
AGRDMAQDTGSGKTRALFLPILOQUMDAASGFSMEQEDALIVAPTRELINQI
YLEAKRPAYGTCVRPVVYGVGTGQIRDLKRCNVLTGTPRLDLMGKRVGLTK
VRYLVLEADRMIDMGPEPMRLVSGPKPSKENVQTLMFSAIPEDIQRLADPLK
TDYFLFLAVGIVGACSDVEQTFVQVTKPAKBSQLIDLKSTGSERTMVEYTRQADP
IATLCOEKPTTSIHGDROMQREOALGFRGKGSVYAVTVGARGIDIPVOYV
NPDLPNNDISYVHRIKRTGRCAGTGAASVYDPAKSELAASLVTLLSTRADQVSWL
BERRAFSGHSISGPNPKRHSATDLKRGQGRFQDGSVTSQLAAGTAADDEME"
ORIGIN
Query Match 27.9%; Score 619.4; DB 5; Length 2058;
Best Local Similarity 66.8%; Pred. No. 3,5e-136;
Matches 882; Conservative 0; Mismatches 438; Indels 0; Gaps 0;
Oy 724 TGAATCAAGAGAGCAAAAGTGAATCAATCAATCCCTCTCTCAAGTGAAGATGAGATC 783
Db 443 TGCAGATGATGAGAGAGGATTAATCAATCGTCCCAACACCTCTGAAGATGAGAGTC 502
Oy 784 CATCTTGCATTAATCAAGAGAGATTAATCAATCAATCAATCAATCAATCAATCAATCA 843
Db 503 CATTTTTCCTCAATGAGTCCGCTGATCAATCAATCAATCAATCAATCAATCAATCAAT 562
Oy 844 AGTGTCTGAGATGAGTCAACCAAGCAATTTGACTTTTGAAGAGCTTAATCTCTGTCA 903
Db 563 TGTCAAGTGAACCAACCAACCAACCAATTAAGATGATGAGAGAGAGAGAGAGAGAG 622
Oy 904 GACACTGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 963
Db 623 GTCCCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 682

```


Db 1775 CACAATCCTGTCCAAAGCCACGACGAGAGTTCCCTCGTGCTAGAGGAGTTGGCTTCAG 1833

Search completed: July 27, 2004, 09:14:36
Job time : 8618 secs

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9585-9590 (2000)
 MEDLINE 20402578
 PUBMED 10920202
 REFERENCE 2 (bases 1 to 2224)
 AUTHORS Castillon, D.H.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-2000) Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA
 FEATURES
 source
 1. 2224
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q11.2-12"
 /tissue_type="testis"
 /dev_stage="adult"
 1. 2224
 /gene="VASA"
 14. 2188
 /gene="VASA"
 /product="DEAD box RNA helicase"
 /protein_id="AF86585.1"
 /db_xref="GI:9438227"
 /translation="MGDEMEAEINPHMSYVPIPEKRYSGENGDNENRTPASSSEM
 DPGSRDRHMKSGPACGRCRPAQRCNKRDNSTMGRCFVGRSGRGRSRRFE
 DGDSGGRWSSSDCDNPNRNKRGSKRGYRDGNNSGCRGRGRGRSGRFGCRGFE
 GLGSFNNDDLPDRCQRTGLFGRSRPVLSGTGNDTSGRSRSGSRGGRYGLNEEV
 ITSGKSNWSENGESSDTQCPKTYLPPEPDEDEISFAHYGTGINFDKXTLIV
 EVSGHDAIPALITFEENALCOLNNNIAAGYKLPVQKSIPIILGRDLMAQOT
 GSGTAFLPLILAHMHGDIPTASRKELOEPECIIVATRELNOIYEAKFSFGT
 CRAVVIYGGTOLGHSIROIVGSCNLTCPGRIMDIKEKXIGKRIKIVYLDEADR
 MLDMGFGRPEMKLISCPMPSKEOROTLMFSAFPEIQRILAEFLKSIYLVPAAGV
 GACRDVQVTVLVGQFSPKREKVELLRINIGDRTVTFETKKADFTPTPLCQKIS
 TSIHDDRERERQALGDFRFGKCPVATSVAAAGLDIENVQVHINIDPSTLIDEX
 VRIIGTRGNGTGRAISPFDESDNHLAQPLVKVLTDAQDQVPMLEIARSTVIGP
 FSGSTRGNVFAVDTRKRSKSTLNTAGFSSSRAPNPVDESDM"

ORIGIN

Query Match 100.0%; Score 2224; DB 9; Length 2224;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTTGAAGTACATGGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 60
 Db 1 ACTTGAAGTACATGGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 60
 QY 61 CTATGTTCCCATATTGAGAGAGATAGCTATTCTGAGAAAATGGAGCAATTTTACG 120
 Db 61 CTATGTTCCCATATTGAGAGAGATAGCTATTCTGAGAAAATGGAGCAATTTTACG 120
 QY 121 GATCCAGCTTCATCATAGAAATGATGAGACCTTCGAGAGAGATCATTTCTGAA 180
 Db 121 GATCCAGCTTCATCATAGAAATGATGAGACCTTCGAGAGAGATCATTTCTGAA 180
 QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAAGAGATGCTGGAGTGATTAAGCG 240
 Db 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAAGAGATGCTGGAGTGATTAAGCG 240
 QY 241 AGATTAATACATCAACATGGGTGTTTGGAGTTGAAAAGAGTTTGGAAAAGAGGTTT 300
 Db 241 AGATTAATACATCAACATGGGTGTTTGGAGTTGAAAAGAGTTTGGAAAAGAGGTTT 300
 QY 301 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGCTTCTGGAGAGATCTAGTATGA 360
 Db 301 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGCTTCTGGAGAGATCTAGTATGA 360
 QY 361 CTGCGAAGATTAATCCACAGGAGGTTTTCACAGAGGCGGCTTATCGAGATGG 420
 Db 361 CTGCGAAGATTAATCCACAGGAGGTTTTCACAGAGGCGGCTTATCGAGATGG 420
 QY 421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGAGGTGAAAGAGTAGTTTCCGAGTTG 480

Db 421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGAGGTGAAAGAGTAGTTTCCGAGTTG 480
 QY 481 CCGTGAAGATTTGGTCTAGGAAGTCCAAATTAATGATAGACCCAGACCAATGATGCA 540
 Db 481 CCGTGAAGATTTGGTCTAGGAAGTCCAAATTAATGATAGACCCAGACCAATGATGCA 540
 QY 541 GCGCATGCTGGGCTTTTGGCTTTAAGAACCAAGTATTAAGTGGCACAGTAATGTTGA 600
 Db 541 GCGCATGCTGGGCTTTTGGCTTTAAGAACCAAGTATTAAGTGGCACAGTAATGTTGA 600
 QY 601 TACTTCTCAAAAGCAGAAAGTGGCAGTGAAGTGAACGAGTGTACAAAGGTTTAAATGA 660
 Db 601 TACTTCTCAAAAGCAGAAAGTGGCAGTGAAGTGAACGAGTGTACAAAGGTTTAAATGA 660
 QY 661 AGAAGTAAATAACAGGCTCTGGAAGAAATTTCTGGAAGTCAAGACGAAAGAGAGAAAG 720
 Db 661 AGAAGTAAATAACAGGCTCTGGAAGAAATTTCTGGAAGTCAAGACGAAAGAGAGAAAG 720
 QY 721 TAGTATACCTCAAGAACCAAAAGTACCTATACCCCTCTCCACCTGAGGATGAGA 780
 Db 721 TAGTATACCTCAAGAACCAAAAGTACCTATACCCCTCTCCACCTGAGGATGAGA 780
 QY 781 CTCCATCTTTCACATTATCAGACAGGCAATAACTTGCACAAATACGACATATTTCTGT 840
 Db 781 CTCCATCTTTCACATTATCAGACAGGCAATAACTTGCACAAATACGACATATTTCTGT 840
 QY 841 GGAAGTCTGGAATATATGACACACAGCAATTTCTGACTTTTGAAGAACCTAATCTCTG 900
 Db 841 GGAAGTCTGGAATATATGACACACAGCAATTTCTGACTTTTGAAGAACCTAATCTCTG 900
 QY 901 TCAGACACTGATATACCAATGCTTAAAGCTGTATTAAGCTTAAGCTTAAGCTTGAAGAA 960
 Db 901 TCAGACACTGATATACCAATGCTTAAAGCTGTATTAAGCTTAAGCTTAAGCTTGAAGAA 960
 QY 961 ATACAGATTTCTTATCATTAATCTGAGAGACAGATTTGATGGCTTGGCTTAAACAGGCTC 1020
 Db 961 ATACAGATTTCTTATCATTAATCTGAGAGACAGATTTGATGGCTTGGCTTAAACAGGCTC 1020
 QY 1021 TGGAGAGCTGCGGCTTTTCTCCATCAATTTTGGCTCATATATGATGATGAAATTAAC 1080
 Db 1021 TGGAGAGCTGCGGCTTTTCTCCATCAATTTTGGCTCATATATGATGATGAAATTAAC 1080
 QY 1081 TGCCAGTCTTTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 Db 1081 TGCCAGTCTTTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 QY 1141 ATTGCTAACAGATTTATTTGAGAGCCAGAAATTTCTTTGGGACTTGTGTAGAGC 1200
 Db 1141 ATTGCTAACAGATTTATTTGAGAGCCAGAAATTTCTTTGGGACTTGTGTAGAGC 1200
 QY 1201 TGTGTTATATATGAGGGAACCCAGCTGGAGATTCATTTGCAAAATAGTACAGGCTG 1260
 Db 1201 TGTGTTATATATGAGGGAACCCAGCTGGAGATTCATTTGCAAAATAGTACAGGCTG 1260
 QY 1261 TAAATATATATGCTACTCTCGGAAGACATGATGATATATAGGCAAAAGAAATGAG 1320
 Db 1261 TAAATATATATGCTACTCTCGGAAGACATGATGATATATAGGCAAAAGAAATGAG 1320
 QY 1321 TCTCAAAACAGATCAAAATAGTTAGTTTGAATGAGAGTATCGCATGTGTGATATGAGTTT 1380
 Db 1321 TCTCAAAACAGATCAAAATAGTTAGTTTGAATGAGAGTATCGCATGTGTGATATGAGTTT 1380
 QY 1381 TGTGTCAGAAATGAAGAGTTAATTTCTTGCACAGATCCATCAAGAGAACAGCGCCA 1440
 Db 1381 TGTGTCAGAAATGAAGAGTTAATTTCTTGCACAGATCCATCAAGAGAACAGCGCCA 1440
 QY 1441 AACCTTAATGTTTCAGTCAACTTTTCAGAGAGAAATCAAGAGTTGCTCAGAGTTT 1500
 Db 1441 AACCTTAATGTTTCAGTCAACTTTTCAGAGAGAAATCAAGAGTTGCTCAGAGTTT 1500
 QY 1501 AAAGTCAATTAATCTGTTTGTGCTGTGAGCAAGTGGTGAAGCATGTAGAGATGTTCA 1560

Db	1501	AAAGTCMAATATATCTGTTTGTTCCTGTTGGACACTGGGTGGAGCATGTAGAGATGTTCA	1560
QY	1561	GCAGACCGTTCCTCCAGTTGCGCAGTTCTCCAAAAGAGAAAAGCTCGTTGAAATCTGCG	1620
Db	1561	GCAGACCGTTCCTCCAGTTGCGCAGTTCTCCAAAAGAGAAAAGCTCGTTGAAATCTGCG	1620
QY	1621	AAACATAGGGGATGGAAGAATCATGTCCTTTGTGAAACTAGAAAAAAGCAGATTTTAC	1680
Db	1621	AAACATAGGGGATGGAAGAATCATGTCCTTTGTGAAACTAGAAAAAAGCAGATTTTAC	1680
QY	1661	TGCAACTTTTCTTTGTCCAGAAAAAAATATATCACTACAAGTATCCATGTGATTCGGGAACA	1740
Db	1661	TGCAACTTTTCTTTGTCCAGAAAAAAATATCACTACAAGTATCCATGTGATTCGGGAACA	1740
QY	1741	GAGAGAGGGGAGAGCAAGCTCTTGGAGATTTTTCGCTTTGGAAAGTCCCAATTTCTTTGTC	1800
Db	1741	GAGAGAGGGGAGAGCAAGCTCTTGGAGATTTTTCGCTTTGGAAAGTCCCAATTTCTTTGTC	1800
QY	1801	TACTTCAGTAGCTGCCAGAGGAGCTGGATATTTGAAATGTCCACAGATTTATCATATTTGA	1860
Db	1801	TACTTCAGTAGCTGCCAGAGGAGCTGGATATTTGAAATGTCCACAGATTTATCATATTTGA	1860
QY	1861	TCTTTCCTTCATCCATGTGATGATATGTTCATTCGAAATTTGGGCGTACTGTGCTGTGTGGAA	1920
Db	1861	TCTTTCCTTCATCCATGTGATGATATGTTCATTCGAAATTTGGGCGTACTGTGCTGTGTGGAA	1920
QY	1921	TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACCATTTAGACAGACGCTCT	1980
Db	1921	TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACCATTTAGACAGACGCTCT	1980
QY	1981	AGTAAAGATATTGACAGATGCTCAACAGGATGTTCTTCGATGCTTTGGAAAGAAATGCGTT	2040
Db	1981	AGTAAAGATATTGACAGATGCTCAACAGGATGTTCTTCGATGCTTTGGAAAGAAATGCGTT	2040
QY	2041	TAGTACATACATTCCTGGCTTCAGTGTGATGACAAAGAGAAACGTGTTGGATCAAGTTGA	2100
Db	2041	TAGTACATACATTCCTGGCTTCAGTGTGATGACAAAGAGAAACGTGTTGGATCAAGTTGA	2100
QY	2101	TACACAGAAAGGGCAAGAGCACTTTGGAACAACAGCTGGGTTTCTTTCCACAGCTCCCAA	2160
Db	2101	TACACAGAAAGGGCAAGAGCACTTTGGAACAACAGCTGGGTTTCTTTCCACAGCTCCCAA	2160
QY	2161	TCCAGTAGATATAGATCATGGAGATTAAGGCCAAACATCCTTCAAGCTGTGTGTTTGA	2220
Db	2161	TCCAGTAGATATAGATCATGGAGATTAAGGCCAAACATCCTTCAAGCTGTGTGTTTGA	2220
QY	2221	TGCA 2224	
Db	2221	TGCA 2224	

RESULT 2					PRI 30-MAY-2000
AF262962	AF262962	2411 bp	mRNA	linear	
LOCUS	Homo sapiens VASA protein mRNA,				
DEFINITION	complete cds.				
ACCESSION	AF262962				
VERSION	AF262962.1	GI:8102020			
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.

REFERENCE	1 (bases 1 to 2411)
AUTHORS	Rocha, D. and Alfara, N.

TITLE Cloning and characterization of the human VASA gene
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2411)
Docha D and Affram M

ADDRESS: ROBERT D. and ALICE, N.
 TITLE: Direct Submission
 FORM NO. 1

Submitted (02-MAY-2000) Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

FEATURES	Location
source	1. .2411

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="5"
18. .2192
CDS

```

```

/codon_start=1
/product="VAsA protein"
/protein_id="AA72705.1"
/db_xref="GI:8102021"
/translation="MGEDWEAEINPMSSGYDPEKORYSGENGDNFRRTPASSSEM
DDGSGSRDHPMKSGFAGRNFGMRDAGECKRNTSMGFGVAGKSGANGFSNRSR
DDGSGSGFRRSSNDCEENPFRNRRGSGYRGGRYVSGNASEAGPYRGGRGSPRCGRG
FLGSPNNDDPDSCMORTGLFGSRRYVSLGTNGDTSQGRSSGSGFRGYKLTNEVY
ITGSGKSMKSSAEAGGSSDTPQGRKTYLPPPEDDTSIFAITYQTEINDKTYITLV
EVSGHDAFPAILFEEANLCQTLNNINAKRGKTLTPVQKSLPIILAGNDACAQT
GSGTAFLFLPILAHMHWDITLASFEELDEPCLIIAPARELVNOYIARAFNSPT
CVAAYVYVGTOGLGSHJROIVOGCNILCAFGELIMDIGEKYGLMOIKYLVDEAR
MMDGFEPMKRLISCGMBSKEOROTIMSAPEPEIGSLAEFLSYLFAVNGV
GACRDVQCYVLQVQSGSKREKYEILRNIGDRTYVATFKKADIAFTLCOEKIS
TTSIHGREOREKQALGDGDFRGKCPVLVATSAAGGLDENVOHINPDLPSSTIDY
VARIARGRCNGTGRAISFPLDSDVHDLAPLKVLTDAQDPAMLEEIAFSTYIPG
FSGSTRNRYASVATDTKRGKSTLNTAGFSSSOANPVDESDMD"

```

ORIGIN

Query Match	90.0%;	Score 2001;	DB 9;	Length 2411;
-------------	--------	-------------	-------	--------------

Best Local Similarity 99.8%; Pred. No. 0;
Matches 2201; Conservative 0; Mismatches 4; Indels 0; Gaps 0

[illegible][illegible]

Db 24 GATGAGATTGGGAGCAGAAATCAACCTCATATGTCTTCCATGTGCCATATTGAG 83

80 AAGTAGTATTCTGGAGAAATGAGACATTTTAACAGGACTCCAGCTTCATCATCA 139

Db 84 AAGCATGCTATTCTGGAGAAATGAGACAAATTTTAACAGGACTCCAGCTTCATCATCA 143

DY 140 GAATGGAATGATTGGACCTTTTGTTCAGAGAATAATCAGAAAGCAATCCATCGCATCGGCCTTG
 168

[illegible]

17 GGGGAGGACCGCCGCGAAGGACATTCATGAAGAAGTAACTGCCTCCTGG 203

200 CGAATTTCGAACAGAGCTGCTGTCAGCTATAAGCAGATATACATCCCAATG 259

Db 204 CGGAATTTGGAAACAGAGATGCTGTGAGTGTAAATAAGCGAGATAAATACATCCACAATG 263

260 GGTGTTTGGAGTTGAAAGAGTTTGGAAACAGAGTTTTCAACAGCAGTTTTGAA 319

Db 264 GGTGTTTGGAGTTGGAAGACTTTTGGAAACAGAGGTTTTCAAACAGCAGGTTTGA 323

320 GATGGTAGCTCTGTTTCTGGAGAGACTAGTATGACTCGAGATATCCACA 379

D5 324 GATGCTGATAGCTCTGGTTTCTGGAGAGAGCTTCTAGTAATGACTCGGAGATTAATCCACA 383

380 CCGAACAAGGGTTTTCCAAAGAGAGCGGCTATCGAGATCGAAATAATTCACAGCCTTCA 430

284

59 CGGACACAGCGGCGGATCGAGATGGAAATATTCGAGGCTTCA 443

440 GGCCATACAGAGAGCTGGAGAGTAGTTCCAGGTTGCCCTGAGAGATTGGTCTA 499

Db 444 GGGCCATACGAGAGGTTGGAGAGGTTAGTTCGAGGTTGCCGTGAGAGATTGTTCTA 503

500 GGAGTCCAATAATGACTTAGACCCAGACGAATGTATGCAGCGCACTGTGGCTTTT 559

Db 504 GGAAGTCCAATAATGACTTTAGACCCAGACGAATGTATGACGGCCACTGTTGGCTTTT 563

560 GGTTCGAGACCACTATTAGTGGACAGCTATGGTGATCTTCTCAAGCAGAGCT 619

Db 564 GGTCTCAAGAGCACCAGTATTTAAGTGGCAGAGGTAATGGTGTACTTCTCAAGCAGAGT 623

620 GGCAGTGGAGTGAACGAGGTGTTACAAAGTTTAAATGACAGAACTAATTAACAGGCTCTT 679

[illegible][illegible]

Db 684 GGAAGAAATCTTGGAAAGTCAGAAAGCAGAAAGGAGAAAGTAGATCTCAAGACCA 743
 Qy 740 AAGTACCTACATACCCCTCTCCACCTGAGAGATGAGATCTCCATCTTTGACATTTAT 799
 Db 744 AAGTACCTACATACCCCTCTCTCCACCTGAGAGATGAGATCTCCATCTTTGACATTTAT 803
 Qy 800 CAGACAGCATTAATCTTGCACAAATACGACATATCTTTGAGAGTGTCTGACATGAT 859
 Db 804 CAGACAGCATTAATCTTGCACAAATACGACATATCTTTGAGAGTGTCTGACATGAT 863
 Qy 860 GACACACAGCATTTCTGACTTTTGAAGAGCTAATCTGTGACAGCATGATTAACAC 919
 Db 864 GACACACAGCATTTCTGACTTTTGAAGAGCTAATCTGTGACAGCATGATTAACAC 923
 Qy 920 ATTGCTAAGCTGTTATTAATTAAGCTTACTCTGTGCAAAAATACAGATTTCTATCA 979
 Db 924 ATTGCTAAGCTGTTATTAATTAAGCTTACTCTGTGCAAAAATACAGATTTCTATCA 983
 Qy 980 CTTCGACAGAGATTTGATGCTGTGTCTCAACAGGCTCTGGAGAGCTCGGCTTTT 1039
 Db 984 CTTCGACAGAGATTTGATGCTGTGTCTCAACAGGCTCTGGAGAGCTCGGCTTTT 1043
 Qy 1040 CTTCCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1099
 Db 1044 CTTCCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1103
 Qy 1100 TTGACAGAACCAAGATGATTTATTTGACACCACTGACAGATTTGATGATGATGATGAT 1159
 Db 1104 TTGACAGAACCAAGATGATTTATTTGACACCACTGACAGATTTGATGATGATGATGAT 1163
 Qy 1160 TTGAGAACCAAGAAATTTCTTTTGGAGCTGTGTGATGATGATGATGATGATGATGATGAT 1219
 Db 1164 TTGAGAACCAAGAAATTTCTTTTGGAGCTGTGTGATGATGATGATGATGATGATGATGAT 1223
 Qy 1220 ACCCAGCTGGACATTTCAATTTGACAAATGATGATGATGATGATGATGATGATGATGAT 1279
 Db 1224 ACCCAGCTGGACATTTCAATTTGACAAATGATGATGATGATGATGATGATGATGATGAT 1283
 Qy 1280 CTTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1339
 Db 1284 CTTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1343
 Qy 1340 TTGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399
 Db 1344 TTGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1403
 Qy 1400 TTGATTTTCTGCCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1459
 Db 1404 TTGATTTTCTGCCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1463
 Qy 1460 ACTTTTCCAGAGAGAAATTTCAAGGTTGCTGACAGATTTTAAAGTCAATATCTGTTT 1519
 Db 1464 ACTTTTCCAGAGAGAAATTTCAAGGTTGCTGACAGATTTTAAAGTCAATATCTGTTT 1523
 Qy 1520 GTTCTGTGTCAGAGAGTGTGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 1579
 Db 1524 GTTCTGTGTCAGAGAGTGTGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 1583
 Qy 1580 GGCAGAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGGAACATATGAGGATGAAAG 1639
 Db 1584 GGCAGAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGGAACATATGAGGATGAAAG 1643
 Qy 1640 ACTATGCTCTTTGTTGAAGTAAAGAAAGAGAGATTTTATGCAACTTTTCTTTGTCAA 1699
 Db 1644 ACTATGCTCTTTGTTGAAGTAAAGAAAGAGAGATTTTATGCAACTTTTCTTTGTCAA 1703
 Qy 1700 GAAAGAAATATCAATTAAGATATCCATGATGATGATGATGATGATGATGATGATGATGAT 1759
 Db 1704 GAAAGAAATATCAATTAAGATATCCATGATGATGATGATGATGATGATGATGATGATGAT 1763
 Qy 1760 CTTCGAGATTTTCTGCTTTGAGAAAGTGCAGATTTCTTTGTCATCTTCAGTAGTGCAG 1819

Db 1764 CTTCGAGATTTTCTGCTTTGAGAAAGTGCAGATTTCTTTGTCATCTTCAGTAGTGCAG 1823
 Qy 1820 GGCCTGATATTTGAAATGTCACATGTTATCAATTTTATCTTCTCTTACATTTGAT 1879
 Db 1824 GGCCTGATATTTGAAATGTCACATGTTATCAATTTTATCTTCTCTTACATTTGAT 1883
 Qy 1880 GAATATGTCATGATTTGGAGCGATGCTGCTGTGTGAGAAATCTGTCAGAGAGATTTTC 1939
 Db 1884 GAATATGTCATGATTTGGAGCGATGCTGCTGTGTGAGAAATCTGTCAGAGAGATTTTC 1943
 Qy 1940 TTTTGTATCTTTGATGATGATTAACATTTAGCAGCTCTAGTAAAGATATGACAGAT 1999
 Db 1944 TTTTGTATCTTTGATGATGATTAACATTTAGCAGCTCTAGTAAAGATATGACAGAT 2003
 Qy 2000 GCTTAACAGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2059
 Db 2004 GCTTAACAGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2063
 Qy 2060 TTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2119
 Db 2064 TTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2123
 Qy 2120 ACTTTGAAACACAGCTGCTTTTCTTCTTCAACAGCTCCCAATCCAGATGATGATGATGAT 2179
 Db 2124 ACTTTGAAACACAGCTGCTTTTCTTCTTCAACAGCTCCCAATCCAGATGATGATGATGAT 2183
 Qy 2180 TGGATTTAAAGCCAAACATCTCTCAAGCTGTGTTGATGCA 2224
 Db 2184 TGGATTTAAAGCCAAACATCTCTCAAGCTGTGTTGATGCA 2228

RESULT 3
 HSM802178
 LOCUS
 DEFINITION Homo sapiens mRNA; cDNA DKFp434B1122 (from clone DKFp434B1122); partial cds.
 ACCESSION AL137462
 VERSION AL137462.1 GI:6808043
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Blum, H., Baereswisch, S., Mewes, H. W., Gassenhuber, J. and Wiemann, S.
 TITLE Direct Submission
 JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

COMMENT
 This clone (DKFp434B1122) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
 Location/Qualifiers

FEATURES
 source
 1..2189
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFp434B1122"
 /tissue_type="testis"
 /clone_id="434 (synonym: htes3). Vector psport1; host DH10B; sites NotI + SalI"
 /dev_stage="adult"
 1..2189
 /gene="DKFp434B1122"
 <1..1910
 /gene="DKFp434B1122"
 /note="strong similarity to testis specific Helicase (VISA)

gene
 CDS

/note="strong similarity to testis specific Helicase (VISA)


```
homolog"/
/codon_start=3
/product="hypothetical protein"
/protein_id="CAB70750.1"
/db_xref="gi:6808044"
/db_xref="GOA:O9NOI0"
/db_xref="Swiss-Prot:O9NOI0"
/translation="SFGRNRSNRFEDSGSRRSSNDCEDEPNRANRERKRGY
RDNNSESGPRRGRGSRFRGCGFGLSPNNDDPDCEMOTGSLFSPSRPVLG
TNGDTSQSRSGSERGKGLNEVITGSGKMSSEAGSSDTPQPKYIYP
PPEDDESIPAHYOTGINFDYDTILVEVGHDPALITFEENLCOITANNIAAG
YTKLPVOKSIPILIAGRDIMACOTGSGKTAFLPLIAPHNNIDITSRFEEIOE
PECTIVATRELINVOIYLEARKEFSFGTCAAVIYGGTOLGHSIROIVOGCNILCAP
GRLMDDIGKEKIKQIKYLVLDABMDLMGRGPKMKLISGQMPSEKRODTLMS
ATPPEIORLAAEFLKSNYLFAVNGVYGACRDYQOTVLOVGOFSREKULVEILRNG
DERTWVETKKADEFIATFLCEKISTISIHDEQERERQALDFRFGCEVLVAT
SVAARGLDIENVQAHINFDLPSTIDEVHAIGRTRGCGRAISFPDLSDNHLAP
LVKVLDAQODVPAMLEIAFSTYIPGFSSTGRNVFASVDTREKKSITNTAGFSSSO
APNVDDESMD"
2128..2133
polyA_signal
/gene="DKFZp434B1122"
2147
polyA_site
/gene="DKFZp434B1122"
```

ORIGIN

Query Match 78.3%; Score 1742; DB 9; Length 2189;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1942; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 279 AGAGTTTGAACAAGAGTTTTCACACAGAGTTTGAAGATGATAGCTCGTT 338
Db 1 AGAGTTTGAACAAGAGTTTTCACACAGAGTTTGAAGATGATAGCTCGTT 60
QY 339 TCTGAGAGAGTCTAGTAATGACTGCCAAGTATTCACACAGGAAGAAGGGTTTCCA 358
Db 61 TCTGAGAGAGTCTAGTAATGACTGCCAAGTATTCACACAGGAAGAAGGGTTTCCA 120
QY 399 AGAAGCGCGCTATCGAGATGGAATATTCAGAACTTCAGGCGCATACAGAAAGAGTG 458
Db 121 AGAAGCGCGCTATCGAGATGGAATATTCAGAACTTCAGGCGCATACAGAAAGAGTG 180
QY 459 GAAAGGTAGTTTCCGAGTTGCCGTGAGAGATTTGGTCTAGAAAGTCCAAATATATCT 518
Db 181 GAAAGGTAGTTTCCGAGTTGCCGTGAGAGATTTGGTCTAGAAAGTCCAAATATATCT 240
QY 519 TAGACCGAGACGATGATGACGCGACTGGTGGCTTTTGGTTCTAGAAAGACAGTAT 578
Db 241 TAGACCGAGACGATGATGACGCGACTGGTGGCTTTTGGTTCTAGAAAGACAGTAT 300
QY 579 TAAGTGGACAGGTATGATGATCTTCTCAAGACAGAGTGGCATGGAAGTGAACGAG 638
Db 301 TAAGTGGACAGGTATGATGATCTTCTCAAGACAGAGTGGCATGGAAGTGAACGAG 360
QY 639 GTGGTTCAAAAGTTTAAATGAAGATTAATCAAGGCTCGGAAATAATCTTGAAGT 698
Db 361 GTGGTTCAAAAGTTTAAATGAAGATTAATCAAGGCTCGGAAATAATCTTGAAGT 420
QY 699 CAGAAGCAGAAAGAGAGAAAGTAGTATCTCAAGACCAAAAGTAGACTACATCCCC 758
Db 421 CAGAAGCAGAAAGAGAGAAAGTAGTATCTCAAGACCAAAAGTAGACTACATCCCC 480
QY 759 CTCTCCACCTGAGAGAGAGACTCCATCTTTGACATTAATCAGACGAGATTAACCTTG 818
Db 481 CTCTCCACCTGAGAGAGAGACTCCATCTTTGACATTAATCAGACGAGATTAACCTTG 540
QY 819 ACAATAACGACACTATTTCTGTGAGAGTGTGACATGATGACACACGACATTTCTGA 878
Db 541 ACAATAACGACACTATTTCTGTGAGAGTGTGACATGATGACACACGACATTTCTGA 600
QY 879 CTTTGAAGAAGACTTAATCTCTGTGACACCTGAATTAACAATTGCTTAAAGCTGGTTATA 938
Db 601 CTTTGAAGAAGACTTAATCTCTGTGACACCTGAATTAACAATTGCTTAAAGCTGGTTATA 660
```

```
QY 939 CTAAAGTTACTCTGTGCAAAAAATACAGTATTCATCATCTTGCAGAGACGAGATTGCA 998
Db 661 CTAAAGTTACTCTGTGCAAAAAATACAGTATTCATCATCTTGCAGAGACGAGATTGCA 720
QY 999 TGGCTGTGCTCAAAACAGGGCTCGGGAAGACTGCCGCTTTTCTCCACCAATTTGGCTC 1058
Db 721 TGGCTGTGCTCAAAACAGGGCTCGGGAAGACTGCCGCTTTTCTCCACCAATTTGGCTC 780
QY 1059 ATATGATGATGATGATGATTAAGTATCTCCAGTGTGTTTAAAGAGTTGACAGAACAGAGTGA 1118
Db 781 ATATGATGATGATGATGATTAAGTATCTCCAGTGTGTTTAAAGAGTTGACAGAACAGAGTGA 840
QY 1119 TTATGTAGCAACCACTCGAGAAATTGTCACACAGATTTATTTGGAAGCCAGAAAAATTTT 1178
Db 841 TTATGTAGCAACCACTCGAGAAATTGTCACACAGATTTATTTGGAAGCCAGAAAAATTTT 900
QY 1179 CTTTGGGACTTGTGTAAAGCTGTGTTATATATGCGGGAACCCAGCTGGGACATTCA 1238
Db 901 CTTTGGGACTTGTGTAAAGCTGTGTTATATATGCGGGAACCCAGCTGGGACATTCA 960
QY 1239 TTGCAAAATAGTACAGAGCTGTATATTTATGTGCTACTCTGGAAGACTGATGATA 1298
Db 961 TTGCAAAATAGTACAGAGCTGTATATTTATGTGCTACTCTGGAAGACTGATGATA 1020
QY 1299 TCATAGGCAAAAGAAAGATTGGTCTCAACAGATCAAAATACTTGTGATGATGAAGCTG 1358
Db 1021 TCATAGGCAAAAGAAAGATTGGTCTCAACAGATCAAAATACTTGTGATGATGAAGCTG 1080
QY 1359 ATGCAATGTTGATATGGTGTGTTGCTCAGAAATGAAGAAGTTAATTTCTGCCAGAA 1418
Db 1081 ATGCAATGTTGATATGGTGTGTTGCTCAGAAATGAAGAAGTTAATTTCTGCCAGAA 1140
QY 1419 TGCATCAAAAGGAACAGCGCCAAACCTTATGTTAGTGCACCTTTCCAGAGGAAATTC 1478
Db 1141 TGCATCAAAAGGAACAGCGCCAAACCTTATGTTAGTGCACCTTTCCAGAGGAAATTC 1200
QY 1479 AAAGTTGGCTCAGAGTTTAAAGTCAAAATATCTGTTGCTGTGAGACAAGTGG 1538
Db 1201 AAAGTTGGCTCAGAGTTTAAAGTCAAAATATCTGTTGCTGTGAGACAAGTGG 1260
QY 1539 GTGAGCATGTGAAGATGTTACGACGACCGTCTCCAACTTGGCCAGTTCTAAAAAGAG 1598
Db 1261 GTGAGCATGTGAAGATGTTACGACGACCGTCTCCAACTTGGCCAGTTCTAAAAAGAG 1320
QY 1599 AAAAGCTGTTGAATTTCTGCAAAACATAGGGGATGAAAGAACTATGCTTTGTTGAAA 1658
Db 1321 AAAAGCTGTTGAATTTCTGCAAAACATAGGGGATGAAAGAACTATGCTTTGTTGAAA 1380
QY 1659 CTAAAGAAAAAGCAGATTTTACTGCACTTTTCTTTGCAAAAAAATATCACTACAA 1718
Db 1381 CTAAAGAAAAAGCAGATTTTATGCACTTTTCTTTGCAAAAAAATATCACTACAA 1440
QY 1719 GTATCATGATGATGAGGAACAGAGAGCGGGAACAAGCTCTTGGAATTTGCGCTTG 1778
Db 1441 GTATCATGATGATGAGGAACAGAGAGCGGGAACAAGCTCTTGGAATTTGCGCTTG 1500
QY 1779 GAAAGTGCCAGTTCTTGTGCTACTTCAATGACTGCAAGAGGCTGATATTTGAAAAATG 1838
Db 1501 GAAAGTGCCAGTTCTTGTGCTACTTCAATGACTGCAAGAGGCTGATATTTGAAAAATG 1560
QY 1839 TCGAACAAGTTATCAATTTTGAATCTTCCCTTCAACATGAGAAATATGTCAGAAATG 1898
Db 1561 TCGAACAAGTTATCAATTTTGAATCTTCCCTTCAACATGAGAAATATGTCAGAAATG 1620
QY 1899 GCGGTACTGCTGTTGTGGAATATCTGCAAGCAATTTCTTTTATCTTGAATCGG 1958
Db 1621 GCGGTACTGCTGTTGTGGAATATCTGCAAGCAATTTCTTTTATCTTGAATCGG 1680
QY 1959 ATTAACATTTAGCAACGCTCTAGTAAAGTATTAACAAGTCTCAACAGATGTTCTTG 2018
Db 1681 ATTAACATTTAGCAACGCTCTAGTAAAGTATTAACAAGTCTCAACAGATGTTCTTG 1740
QY 2019 CATGTGGAAGAAATGTGCTTATGATACATATCTGCGCTTCAAGTGTAGTACAAAG 2078
```

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/JMIL at: <http://image.lnl.gov>

Series: IRAX Plate: 91 Row: A Column: 10

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9507236.

```

FEATURES             source
    source            1..2363
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="MGC:48304 IMAGE:5298807"
                        /tissue_type="Testis"
                        /clone_id="NH MGC_97"
                        /lab_host="DH10B"
                        /note="vector: pBluescript"
    gene              1..2363
                        /gene="VDX4"
                        /note="synonym: VASA"
                        /db_xref="LocusID:54514"
                        /db_xref="MIM:605281"
    CDS                38..2110
                        /codon_start=1
                        /product="VDX4 protein"
                        /protein_id="AAH47455.1"
                        /db_xref="GI:28958131"
                        /translation="MGDEWAEAINPMSLYDIPEKORYSGENDGNFNPASSSEM
                        DDPGRDRDHFMKGSPASGRNFGADGECKRNTSMGFGYGSXGNFSRSRE
                        DGSSGWRSSNDCEHNPIRNNGSRGCGNDIDPCSCORTGLFESRRPVLTGDN
                        GDTSQBSGSGSERGGYKGLNEBYITSGSKMSKEBAEGESSDTPCYTYTIPPPP
                        EDDSDSPAHVQTGINFDKYDTLLVESGHAPPAIIFEEANLCOTINNIYAAGYTK
                        LTFVOKYSIPRIILAIRIGACQAQSGETKAFLPLIAHWMDGITASRKEIQEPKC
                        ILVAPELKNYLIYLEARKPSFGTCVAAYIVYGTOGHSIRQIVOGCNICATPGRL
                        MDIIGKEIKGKOIRKYLVEADRMLDMGSPENMKLISCPGMPSKEOROTLMFSATFE
                        PEELORIAAEFKSNYLFVAVGOGGACRVOCOTVLQGVGSKREKLVELIRINDGBR
                        TMVFVERKKADRIATATLCOEKISTSHGDREOREEOLGPERRKPCVIVATSVAFA
                        ARGEDINVOHVINFDFDLSTIDEVHRIGTRCQRGTALSPFDLESNHLAOPLVAK
                        VLTDQQDVPAWLEELAFSTYIPFGSGSTRGNVASFVKRKSITLTAFGSSOAPN
                        PVDSEMD"
                        716..1915
                        /note="SmbB; Region: Superfamily II DNA and RNA helicases
                        /name="SmbB; Region: Superfamily II DNA and RNA helicases
                        /function="DNA replication, recombination, and repair /
                        Transcription / Translation, ribosomal structure and
                        biogenesis"]
                        /db_xref="CCD:COG0513"
    misc_feature      63.4%; Score 1409; DB 9; Length 2363;
    Query Match       Best Local Similarity 99.7%; Pred. No. 0;
                       Matches 1709; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
    ORIGIN
        Query Match   63.4%; Score 1409; DB 9; Length 2363;
        Best Local Similarity 99.7%; Pred. No. 0;
        Matches 1709; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
        510 ATATGACTTAGACCAGACCAATGTATGACGCGACTGTGCTTTGGTTCTTGAA 569
        Db            432 ATATGACTTAGACCAGACCAATGTATGACGCGACTGTGCTTTGGTTCTTGAA 491
        Qy           570 GACCAATTAAGTGGACAGCTTAATGCTATTCTTCAAAGAGAAGTGGCATGGA 629
        Db           492 GACCAATTAAGTGGACAGCTATGATGATACTTCTCAAAGAGAAGTGGCATGGA 551
        Qy          630 GTGAACAGAGTGTTCAAAGGTTTAATGAAGAGTAAATACAGGCTCTGGAAGAATT 689
        Db          552 GTGAACAGAGTGTTCAAAGGTTTAATGAAGAGTAAATACAGGCTCTGGAAGAATT 611
        Qy          690 CTTCGAAGTCAAGAACGAAAGACAGAGAAAGTAGTATCTTCAAGAGCCAAAAGTAGACT 749
        Db          612 CTTCGAAGTCAAGAACGAAAGACAGAGAAAGTAGTATCTTCAAGAGCCAAAAGTAGACT 671
        Qy          750 ACATACCCTCTCCACCTGAGGATAGAGATCCATCTTGCACATTATCAGACAGCA 809
        Db          672 ACATACCCTCTCCACCTGAGGATAGAGATCCATCTTGCACATTATCAGACAGCA 731

```


Db 532 TATCACTTGCAGAGAGATTGATGGCTTGCCTCAAGGCTCTGGAGAGCTGC 591
 Qy 1033 GCGCTTTCTCTACCAATTTTGGCTCATATGATGATGAAATACCTGCCAGTCGTT 1092
 Db 592 GCGTTTCTCTCAATTTTGGCTCATATGATGATGAAATACCTGCCAGTCGTT 651
 Qy 1093 TAAAGATTGCAGAAACAGAGTGTATTTATTTAGACCACTGAGAAATGGTCAACA 1152
 Db 652 TAAAGATTGCAGAAACAGAGTGTATTTATTTAGACCACTGAGAAATGGTCAACA 711
 Qy 1153 GATTTATTTGAAAGCCAGAAATTTTCTTTTGGGACTTGTGTAAAGCTGTGTATTA 1212
 Db 712 GATTTATTTGAAAGCCAGAAATTTTCTTTTGGGACTTGTGTAAAGCTGTGTATTA 771
 Qy 1213 TGGGGGAAACCACTGGGACATTCATTCAGCAATAGTACAAAGCTGTATATATATG 1272
 Db 772 TGGGGGAAACCACTGGGACATTCATTCAGCAATAGTACAAAGCTGTATATATG 831
 Qy 1273 TGGTCTCTGGAAGACTGATGATATCATAGCCAAAGAAAGATTTGTCTCAACAGAT 1332
 Db 832 TGGTCTCTGGAAGACTGATGATATCATAGCCAAAGAAAGATTTGTCTCAACAGAT 891
 Qy 1333 CAATATCTTGTGTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392
 Db 892 CAATATCTTGTGTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
 Qy 1393 GAAAGATTAAATTTCTGCCCAGAGAAATGCCATCAAGAGAAACAGCCCAACCTTATGTT 1452
 Db 952 GAAAGATTAAATTTCTGCCCAGAGAAATGCCATCAAGAGAAACAGCCCAACCTTATGTT 1011
 Qy 1453 CAGTGCACCTTTTCCAGAGAAATTCAGAGTTGGCTGAGAGTTTAAAGTCAATTA 1512
 Db 1012 CAGTGCACCTTTTCCAGAGAAATTCAGAGTTGGCTGAGAGTTTAAAGTCAATTA 1071
 Qy 1513 TCTGTTGTTGCTGTTGGAACAAGTGGTGGAGCATGTAGAGATGTCAGCAGACGTTCT 1572
 Db 1072 TCTGTTGTTGCTGTTGGAACAAGTGGTGGAGCATGTAGAGATGTCAGCAGACGTTCT 1131
 Qy 1573 CCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTTCGTTGAATTTCTGCGAAACATAGGGA 1632
 Db 1132 CCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTTCGTTGAATTTCTGCGAAACATAGGGA 1191
 Qy 1633 TGAAGAACTATGTTCTTTGTAAGTAAAGAGAGATTTTACTGCAACTTTTCT 1692
 Db 1192 TGAAGAACTATGTTCTTTGTAAGTAAAGAGAGATTTTACTGCAACTTTTCT 1251
 Qy 1693 TTGTCAAGAAAATATCAATCAAGTATCATGTTGATGCGGAAACAGAGAGCGGA 1752
 Db 1252 TTGTCAAGAAAATATCAATCAAGTATCATGTTGATGCGGAAACAGAGAGCGGA 1311
 Qy 1753 GCAAGCTCTGAGATTTTGGCTTGGAAAGTCCAGTTCTGTTGCTACTTCACTAGTAC 1812
 Db 1312 GCAAGCTCTGAGATTTTGGCTTGGAAAGTCCAGTTCTGTTGCTACTTCACTAGTAC 1371
 Qy 1813 TGGCAGAGGCTGGATTTGAAAATGTCAAATGTTATCATTTTATCTTCTCTAC 1872
 Db 1372 TGGCAGAGGCTGGATTTGAAAATGTCAAATGTTATCATTTTATCTTCTCTAC 1431
 Qy 1873 CATTTAGATATGTTATGATTTGGGCTACTGCTGTGTGGAAATATCTGGCAGAGC 1932
 Db 1432 CATTTAGATATGTTATGATTTGGGCTACTGCTGTGTGGAAATATCTGGCAGAGC 1491
 Qy 1933 AATTTCTTTTGTATTTGATTCGATTAACATTTAGCAGACCTCTAGTAAAGTAT 1992
 Db 1492 AATTTCTTTTGTATTTGATTCGATTAACATTTAGCAGACCTCTAGTAAAGTAT 1551
 Qy 1993 GACAGATGCTCAACAGAGATTTCTGCTGATGTTGGAAGAAATGCTTTAGTACATAC 2052
 Db 1552 GACAGATGCTCAACAGAGATTTCTGCTGATGTTGGAAGAAATGCTTTAGTACATAC 1611
 Qy 2053 TCCGAGCTTCAAGTGTAGTCAAGAGAAAGCTGTTGCTAGTGTATACAGAGAGG 2112
 Db 1612 TCCGAGCTTCAAGTGTAGTCAAGAGAAAGCTGTTGCTAGTGTATACAGAGAGG 1671

Qy 2113 CAAGAGCACTTGAACACAGCTGGGTTTCTTCTTCTTCAAGAGCTCCCATTCAGTATGTA 2172
 Db 1672 CAAGAGCACTTGAACACAGCTGGGTTTCTTCTTCTTCAAGAGCTCCCATTCAGTATGTA 1731
 Qy 2173 TGAATCATGGGATTAAAGCCAAACATCTTCAAGTGTGTGTTTGAATGCA 2224
 Db 1732 TGAATCATGGGATTAAAGCCAAACATCTTCAAGTGTGTGTTTGAATGCA 1783
 RESULT 6
 AK093439
 LOCUS
 DEFINITION
 Homo sapiens cDNA FLJ36120 fls, clone TEST12022738, highly similar
 to DEAD BOX PROTEIN 4.
 AK093439
 ACCESSION
 AK093439.1 GI:21752310
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
 Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Makamatsu,A., Ishii,S.,
 Yamamoto,J., Isono,Y., Kawai,Hio,Y., Saito,K., Nishikawa,T.,
 Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
 Kikuchi,H., Kanda,K., Wagatsuma,M., Mukakawa,K., Kanehori,K.,
 Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
 Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.
 NEDO human cDNA sequencing project
 TITLE
 JOURNAL
 UNPUBLISHED
 2 (bases 1 to 1984)
 REFERENCE
 AUTHORS
 Isogai,T. and Yamamoto,J.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 source
 1. 1984
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="TEST12022738"
 /issue_type="testis"
 /clone_lib="TEST12"
 /note="Cloning vector: pMB18SFL3"
 ORIGIN
 Query Match 57.9%; Score 1288; DB 9; Length 1984;
 Best local similarity 99.7%; Pred. No. 0;
 Matches 1488; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 733 AGAAGCAAAAGTACCTTACATACCCCTCTCTCACTGAGATGAGAGCTTCATCTTTGC 792
 Db 292 AGAAGCAAAAGTACCTTACATACCCCTCTCTCACTGAGATGAGAGCTTCATCTTTGC 351
 Qy 793 ACATTATGACAGAGCACTTAACTTGAACAATACGACATTTCTTGTGGAAGTGTGCG 852
 Db 352 ACATTATGACAGAGCACTTAACTTGAACAATACGACATTTCTTGTGGAAGTGTGCG 411
 Qy 853 ACATGATCACACAGAGCACTTGTACTTTTGAAGAGCTATCTGTGTACAGACATGAA 912
 Db 412 ACATGATCACACAGAGCACTTGTACTTTTGAAGAGCTATCTGTGTGTGACACATGAA 471

QY	913	TAAACAACATGCTAAAGCTGGTTACTACTAAGCTTACTCCTGGACAAAATACAGTAATCC	972
Db	472	TAAACAACATGCTAAAGCTGGTTACTACTAAGCTTACTCCTGGACAAAATACAGTAATCC	531
QY	973	TATCACTACTGACAGACAGATTTGATGGCTTGCTCAACAGGGCTCTGGAGACATGC	1032
Db	532	TATCACTACTGACAGACAGATTTGATGGCTTGCTCAACAGGGCTCTGGAGACATGC	591
QY	1093	GGCTTTCTCCACCAATTTTGGCTCATATGATGATATGGAATTAACATGCACAGCTTT	1092
Db	592	GGCTTTCTCCACCAATTTTGGCTCATATGATGATGATGGAATTAACATGCACAGCTTT	651
QY	1093	TAAAGATGACAGAAACAGAGTGTATTTATTTGTACCAACCACTCGAATTTGTCACACA	1152
Db	652	TAAAGATGACAGAAACAGAGTGTATTTATTTGTACCAACCACTCGAATTTGTCACACA	711
QY	1153	GATTTATTTGGAAGCCAGAAAATTTCTTTTGGACCTTGTTAAAGCTGTTGTATATA	1212
Db	712	GATTTATTTGGAAGCCAGAAAATTTCTTTTGGACCTTGTTAAAGCTGTTGTATATA	771
QY	1213	TGGGGGAACCCAGCTGGGACATTCATTCGACAAATATATACAAAGCTGTAAATATATAT	1272
Db	772	TGGGGGAACCCAGCTGGGACATTCATTCGACAAATATATACAAAGCTGTAAATATATAT	831
QY	1273	TGCTACTCTTGGAAGACTGATGATATCATATAGCAAGAAAAGATTGGTCTTCAAACAGAT	1332
Db	832	TGCTACTCTTGGAAGACTGATGATATCATATAGCAAGAAAAGATTGGTCTTCAAACAGAT	891
QY	1333	CAATTACTTAATTTTGGATGGAAGCTGATCGCATGTGTGAATATGGTTTTGGCCAGAAAT	1392
Db	892	CAATTACTTAATTTTGGATGGAAGCTGATCGCATGTGTGAATATGGTTTTGGCCAGAAAT	951
QY	1393	GAAGAAGTTAATTTCTTGCCAGGAATCCATCAAAAGGAACAGCCCAACCCCTATATGT	1452
Db	952	GAAGAAGTTAATTTCTTGCCAGGAATCCATCAAAAGGAACAGCCCAACCCCTATATGT	1011
QY	1453	CAGTGCACCTTTTCCAGAGGAATTCAAAGTTGGCTGACAGATTTTAAAGTCAAAATTA	1512
Db	1012	CAGTGCACCTTTTCCAGAGGAATTCAAAGTTGGCTGACAGATTTTAAAGTCAAAATTA	1071
QY	1513	TCTGTTTTGTGCTGTGGACAAGTGGGTGAGACATGTAGAGATGTTCACGACACCGTCT	1572
Db	1072	TCTGTTTTGTGCTGTGGACAAGTGGGTGAGACATGTAGAGATGTTCACGACACCGTCT	1131
QY	1573	CCAAAGTGGCCAGATTCCTAAAAAAGAGAAAGCTGTTGAATTTCTGCGAAAATATAGGGGA	1632
Db	1132	CCAAAGTGGCCAGATTCCTAAAAAAGAGAAAGCTGTTGAATTTCTGCGAAAATATAGGGGA	1191
QY	1633	TGAAGAACAATATGCTTTGTGAAAACATTAAGAAAAACACAGTTTATATGCAACCTTTCT	1692
Db	1192	TGAAGAACAATATGCTTTGTGAAAACATTAAGAAAAACACAGTTTATATGCAACCTTTCT	1251
QY	1693	TTGTCAAGAAAAAATATCACTACACATATCCATGGTATCGGAAACAGAGAGACGGGA	1752
Db	1252	TTGTCAAGAAAAAATATCACTACACATATCCATGGTATCGGAAACAGAGAGACGGGA	1311
QY	1753	GCAAGCTCTTGGAAATTTTGGCTTTGGAAATGAGCCCACTTTGTGTGCTACTCATATAGC	1812
Db	1312	GCAAGCTCTTGGAAATTTTGGCTTTGGAAATGAGCCCACTTTGTGTGCTACTCATATAGC	1371
QY	1813	TGCACAGAGGGCTGATATTTGAAAAATGTGCAACATGTTATCAATTTTGAATCTTCTCTAC	1872
Db	1372	TGCACAGAGGGCTGATATTTGAAAAATGTGCAACATGTTATCAATTTTGAATCTTCTCTAC	1431
QY	1873	CATTGATGAATATGTTCAATCGAATTTGGCGTATCTGCTGTGTGGGAATATCTGCAAGGC	1932
Db	1432	CATTGATGAATATGTTCAATCGAATTTGGCGTATCTGCTGTGTGGGAATATCTGCAAGGC	1491
QY	1933	AATTTCCTTTTTATCTTTGAATCGGAATACATTTATACACAGCCTCTAGATGAAGATTT	1992
Db	1492	AATTTCCTTTTTATCTTTGAATCGGAATACATTTATACACAGCCTCTAGATGAAGATTT	1551
QY	1993	GACAGATGCTCAACAGATATTTCTGTCAGTGTGGAGAAAATTGCTTTTAACTACATACAT	2052

Db	1552	GACGAGTCTCAACAGGATGTTCTCGATGGTTGGAGAAATTGCTTTAGTACATCAT	1611
Qy	2053	TTCTGGCTTCAGTGTGTGTGTACAGAGGAAACGTGTGGTTCATCAGTTGTATACAGAAAGG	2112
Db	1612	TTCTGGCTTCAGTGTGTGTGTACAGAGGAAACGTGTGGTTCATCAGTTGTATACAGAAAGG	1671
Qy	2113	CAAGAGCACTTGAACACAGCTGGGTTTTCTTCTCACAGAGCTCCCAATCAGTAGATGA	2172
Db	1672	CAAGAGCACTTGAACACAGCTGGGTTTTCTTCTCACAGAGCTCCCAATCAGTAGATGA	1731
Qy	2173	TGAGTCATGGGATTAAAGCCAAACATCTCTTCAAGTCTGTGGTTTGATGCA	2224
Db	1732	TGAGTCATGGGATTAAAGCCAAACATCTCTTCAAGTCTGTGGTTTGATGCA	1783
RESULT 7			
LOCUS	AR415286	504 bp	DNA
DEFINITION	Sequence 2923 from patent US 6639063.	linear	PAT 18-DEC-2003
ACCESSION	AR415286		
VERSION	AR415286.1	GI:40170396	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 504)		
AUTHORS	Edwards,J.-B.D.M., Jobert,S. and Giordano,J.-Y.		
TITLE	EST's and encoded human proteins		
JOURNAL	Patent: US 6639063-A,2923 28-Oct-2003;		
FEATURES	Location/Qualifiers		
source	1..504		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	20.1%; Score 448; DB 6; Length 504;		
Best Local Similarity	100.0%; Pred. No. 4,6e-237;		
Matches 448; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	775	TGAGGATCCATCTTTGACATTATCGACAGGATAACTTGACAAATACGACATAT	834
Db	1	TGAGGATCCATCTTTGACATTATCGACAGGATAACTTGACAAATACGACATAT	60
Qy	835	TCCTTGGGAAGTCTTGGACATGATGCAACACGACATTTCTGACTTTGGAGAAGCTAA	894
Db	61	TCCTTGGGAAGTCTTGGACATGATGCAACACGACATTTCTGACTTTGGAGAAGCTAA	120
Qy	895	TCCTGTCCAGACACTGAATTAACAACATTTGCTAAAGCTGTATCTAAGCTTACTCTGT	954
Db	121	TCCTGTCCAGACACTGAATTAACAACATTTGCTAAAGCTGTATCTAAGCTTACTCTGT	180
Qy	955	GCAAAAATACAGTATTCCTATCATACTTGCAGGACGAGATTTGATGGCTTGTGCTCAAC	1014
REFERENCE	181	GCAAAAATACAGTATTCCTATCATACTTGCAGGACGAGATTTGATGGCTTGTGCTCAAC	240
Qy	1015	AGGCTCTGGGAACACTGGCGCTTTTCTCTCAACCAATTTTGGCTCATATGATGATATGG	1074
Db	241	AGGCTCTGGGAACACTGGCGCTTTTCTCTCAACCAATTTTGGCTCATATGATGATATGG	300
Qy	1075	AATTAAGTCCAGTGTGTTTAAAGATTTGCAGGAACCGAGTGTATTTGTAGACCAAC	1134
Db	301	AATTAAGTCCAGTGTGTTTAAAGATTTGCAGGAACCGAGTGTATTTGTAGACCAAC	360
Qy	1135	TCGAGATTTGTCAACCAATTTATTTTGGAAAGCCAGAAAATTTTCTTTGGGACTTGTGT	1194
Db	361	TCGAGATTTGTCAACCAATTTATTTTGGAAAGCCAGAAAATTTTCTTTGGGACTTGTGT	420
Qy	1195	AAGAGCTGTTTATATATATGGGGAAC	1222
Db	421	AAGAGCTGTTTATATATATGGGGAAC	448

RESULT 8
BD110839 504 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION EST and encoded human protein.
ACCESSION BD110839
VERSION BD110839.1 GI:23205657
KEYWORDS JP 2002010789-A/2916.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 504)
Edwards, J.B.D.W., Robert, S. and Giordano, J.E.
EST and encoded human protein
Patent: JP 2002010789-A 2916 15-JAN-2002;
JOURNAL GENSER CORP
COMMENT OS Homo sapiens (human)
PN JP 2002010789-A/2916
PD 15-JAN-2002
PF 07-AUG-2000 JP 2002080989
PR 05-AUG-1999 US 60/147499
PI JEAN BAPTIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FT Location/Qualifiers
FH Key 224.472.
FI CDS
FEATURES
source
1..504
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 20.1%; Score 448; DB 6; Length 504;
Best Local Similarity 100.0%; Pred. No. 4,6e-237;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 775 TGAGAGCTCATCTTTGACATATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 834
DB 1 TGAGAGCTCATCTTTGACATATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 60
QY 835 TCTTGTGAGAGTGTGAGATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 894
DB 61 TCTTGTGAGAGTGTGAGATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 120
QY 895 TCTTGTGAGAGTGTGAGATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 954
DB 121 TCTTGTGAGAGTGTGAGATATGACAGGAGCATTAAGTTCGACAAATACGACATAT 180
QY 955 GCAAAATACAGTATCTTCTATCTATCTGAGAGAGATTTGATGGCTTGTCTAAAC 1014
DB 181 GCAAAATACAGTATCTTCTATCTATCTGAGAGAGATTTGATGGCTTGTCTAAAC 240
QY 1015 AGGCTCGGAGAGAGTCTGGCTTCTTCTACCAATTTGGCTCATATGATGATGATG 1074
DB 241 AGGCTCGGAGAGAGTCTGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300
QY 1075 AATTAAGTGTGATTTTAAAGATTTGAGAGAGATTTGATGGCTTGTCTAAAC 1134
DB 301 AATTAAGTGTGATTTTAAAGATTTGAGAGAGATTTGATGGCTTGTCTAAAC 360
QY 1135 TCGAGATTTGATTTTAAAGATTTGAGAGAGATTTTCTTTTGGAGCTTGTGT 1194
DB 361 TCGAGATTTGATTTTAAAGATTTGAGAGAGATTTTCTTTTGGAGCTTGTGT 420
QY 1195 AAGAGCTGTGTATATATATGAGAGAGCT 1222
DB 421 AAGAGCTGTGTATATATATGAGAGAGCT 448

RESULT 9
AC008914 139677 bp DNA linear PRI 19-APR-2002
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2270116, complete sequence.
ACCESSION AC008914
VERSION AC008914.7 GI:20198513
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 139677)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE
2 (bases 1 to 139677)
DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE
3 (bases 1 to 139677)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE
4 (bases 1 to 139677)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE
Submitted (19-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 19, 2002 this sequence version replaced gi:19747123.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1..139677
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2270116"
ORIGIN
Query Match 12.3%; Score 273; DB 9; Length 139677;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1728 GTGATCGGGAACAGAGAGAGCGGAGCAAGCTTTGAGATTTTGGTGGAAAGTGC 1787
DB 13765 GTGATCGGGAACAGAGAGAGCGGAGCAAGCTTTGAGATTTTGGTGGAAAGTGC 13824
QY 1788 CAGTTCTTTGCTACTTCACTAGCTGCGAGAGGCTGATATTTGAAAATGTGAAATG 1847
DB 13825 CAGTTCTTTGCTACTTCACTAGCTGCGAGAGGCTGATATTTGAAAATGTGAAATG 13884
QY 1848 TTATCAATTTGATCTTCTTCTTCACTTGAATGATGATGATGATGATGATGATGATG 1907
DB 13885 TTATCAATTTGATCTTCTTCTTCACTTGAATGATGATGATGATGATGATGATG 13944
QY 1908 GTGCTGTGGGAATACGAGAGCAATTTCTTTTGTATCTTGAATCGGATTAACATT 1967
DB 13945 GTGCTGTGGGAATACGAGAGCAATTTCTTTTGTATCTTGAATCGGATTAACATT 14004
QY 1968 TAGCAGAGCTCTAGTAAAGATTTGACAGATG 2000
DB 14005 TAGCAGAGCTCTAGTAAAGATTTGACAGATG 14037

RESULT 10

AC016632/C	AC016632	177185 bp	DNA	linear	PRI 01-DEC-2002
LOCUS	AC016632	177185 bp	DNA	linear	PRI 01-DEC-2002
DEFINITION	Homo sapiens chromosome 5 clone RP11-175M2, complete sequence.				
ACCESSION	AC016632				
VERSION	AC016632.7	GI:25989044			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 177185)				
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 177185)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
REFERENCE	3 (bases 1 to 177185)				
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
COMMENT	On Dec 1, 2002 this sequence version replaced gi:15290348.				
	Draft Sequence Produced by DOE Joint Genome Institute				
	www.jgi.doe.gov				
	Finishing Completed at Stanford Human Genome Center				
	www.shgc.stanford.edu				
	Quality: Phrap Quality >=40 99.4% of Sequence;				
	Estimated Total Number of Errors is 0.7.				
FEATURES	Location/Qualifiers				
source	1..177185				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="5"				
	/clone="RP11-175M2"				
ORIGIN					
Query Match	12.3%; Score 273; DB 9; Length 177185;				
Best Local Similarity	100.0%; Pred. No. 9.3e-140;				
Matches 273; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1728 GTGATCGGAAACAAGAGAGGCGGAGCAAGCTCTTGAGATTTCCGCTTTGGAAAGTGCC	1787			
Db	34889 GTGATCGGAAACAAGAGAGGCGGAGCAAGCTCTTGAGATTTCCGCTTTGGAAAGTGCC	34830			
QY	1788 CAGTCTTGTTGCTACTTCAGTAGCTCCGAGAGGGCTGAGATTGTAAGATGTCACATG	1847			
Db	34823 CAGTCTTGTTGCTACTTCAGTAGCTCCGAGAGGGCTGAGATTGTAAGATGTCACATG	34770			
QY	1848 TTATCAATTTGATCTCTTCCTTACCAATTGATGATATGTTCAATGCAATTGGCGCTAGT	1907			
Db	34769 TTATCAATTTGATCTCTTCCTTACCAATTGATGATATGTTCAATGCAATTGGCGCTAGT	34710			
QY	1908 GTCGTTGGGGAATACAGGAGAGCAATTTCTTTTGTCTTGTAATGGATAACCAATT	1967			
Db	34709 GTCGTTGGGGAATACAGGAGAGCAATTTCTTTTGTCTTGTAATGGATAACCAATT	34650			
QY	1968 TAGCACAGCCTCTAGTAAAGATATTGACAGATG	2000			
Db	34649 TAGCACAGCCTCTAGTAAAGATATTGACAGATG	34617			
RESULT 11					
AC016639/C	AC016639	182126 bp	DNA	linear	HTG 18-JUL-2000
DEFINITION	Homo sapiens chromosome 5 clone RP11-412L4, WORKING DRAFT SEQUENCE,				
	7 ordered pieces.				
ACCESSION	AC016639				
VERSION	AC016639.5	GI:7711585			
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.				

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE   1 (bases 1 to 182126)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 182126)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On May 6, 2000 this sequence version replaced gi:1770162.

COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

-----
Project Information
Center Project Name: 56481
Center clone name: RPCI-11_412L4
-----

Summary Statistics
Consensus quality: 175214 bases at least Q40
Consensus quality: 180404 bases at least Q30
Consensus quality: 181112 bases at least Q20
Estimated insert size: 178000; pulse field gel estimation
Estimated insert size: 181876; sum-of-contigs estimation
Quality coverage: 6.41 in Q20 bases; pulse field gel estimation
Quality coverage: 6.28 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1      48575: contig of 48575 bp in length
*      48676: contig of 14135 bp in length
*      62811: gap of unknown length
*      62911: gap of unknown length
*      81913: contig of 19003 bp in length
*      81914: gap of unknown length
*      82014: contig of 10153 bp in length
*      92167: gap of unknown length
*      92266: gap of unknown length
*      92267: contig of 75212 bp in length
*      167479: gap of unknown length
*      167578: gap of unknown length
*      167579: gap of unknown length
*      178445: contig of 10867 bp in length
*      178446: gap of unknown length
*      178545: gap of unknown length
*      178546: contig of 3581 bp in length.

Location/Qualifiers
1. 182126
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-412L4"
/clone_11b="RPCI human BAC library 11"

ORIGIN
Query Match      12.3%; Score 273; DB 2; Length 182126;
Best Local Similarity 100.0%; P-Id No. 9,3e-140;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1728 GTGATCGGGAACAGAGAGACGGGAGAGACCTTTGGAGATTTCGCTTGAAGAGGCC 1787
Db      36792 GTGATCGGGAACAGAGAGACGGGAGAGACCTTTGGAGATTTCGCTTGAAGAGGCC 3673

OY      1788 CAGTTCTTTGCTACTTCACTAGTACCTGACAGAGGCGCTGATATTGAAATGTCACATG 1847

```

```

Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
-----
Project Information -----
Center project name: H_NH033C17
-----
Summary Statistics -----
Sequencing vector: M13, 87%
Sequencing vector: plasmid, 13%
Chemistry: Dye-terminator Big Dye, 15% of reads
Chemistry: Dye-terminator Big Dye, 15% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 196459 bases at least Q40
Consensus quality: 199226 bases at least Q30
Consensus quality: 200804 bases at least Q20
Insert size: 205000, agarose-fp
Insert size: 202150, sum-of-contigs
Quality coverage: 4.50 in Q20 bases, agarose-fp
Quality coverage: 4.65 in Q20 bases, sum-of-contigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
1 1062: contig of 1062 bp in length
1063 1162: gap of unknown length
1163 2536: contig of 1374 bp in length
2537 2636: gap of unknown length
2637 5791: contig of 3135 bp in length
5792 5792: gap of unknown length

```

	*	5892	10946:	contig of 5055 bp in length
	*	10947	11046:	gap of unknown length
	*	11047	15725:	contig of 4679 bp in length
	*	15726	15825:	gap of unknown length
	*	15826	21169:	contig of 5344 bp in length
	*	21170	21269:	gap of unknown length
	*	21270	27800:	contig of 6531 bp in length
	*	27801	27900:	gap of unknown length
	*	33277	33276:	contig of 5376 bp in length
	*	33377	42711:	gap of unknown length
	*	42712	42811:	gap of unknown length
	*	42812	56046:	contig of 13235 bp in length
	*	56047	56146:	gap of unknown length
	*	56147	77263:	contig of 21117 bp in length
	*	77264	77363:	gap of unknown length
	*	77364	97974:	contig of 20611 bp in length
	*	97975	98074:	gap of unknown length
	*	98075	122272:	contig of 24198 bp in length
	*	122273	122372:	gap of unknown length
	*	122373	148199:	contig of 25827 bp in length
	*	148200	148299:	gap of unknown length
	*	148300	175035:	contig of 26736 bp in length
	*	175036	175135:	gap of unknown length
	*	175136	204250:	contig of 29115 bp in length.
FEATURES				
source				
			1.	204250
			/organism="Homo sapiens"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:9606"	
			/chromosome="5"	
			/clone="RP11-332C17"	
			1.	1062
misc_feature			/note="assembly_name:Contig12"	
			1163.	12536
misc_feature			/note="assembly_name:Contig13"	
			2637.	15791
misc_feature			/note="assembly_name:Contig14"	
			5892.	110946
misc_feature			/note="assembly_name:Contig15"	
			11047.	15725
misc_feature			/note="assembly_name:Contig16"	
			15826.	21169
misc_feature			/note="assembly_name:Contig17"	
			21270.	27800
misc_feature			/note="assembly_name:Contig18"	
			27901.	33276
misc_feature			/note="assembly_name:Contig19"	
			33377.	42711
misc_feature			/note="assembly_name:Contig20	
			clone_end:T7	
			vector_side:right"	
			42812.	36046
misc_feature			/note="assembly_name:Contig21"	
			56147.	77263
misc_feature			/note="assembly_name:Contig22"	
			77364.	97974
misc_feature			/note="assembly_name:Contig23"	
			98075.	122272
misc_feature			/note="assembly_name:Contig24"	
			122373.	148199
misc_feature			/note="assembly_name:Contig25"	
			148300.	175035
misc_feature			/note="assembly_name:Contig26"	
			175136.	204250
misc_feature			/note="assembly_name:Contig27"	
ORIGIN				
Query Match			12.3%;	Score 273; DB 2; Length 204250;
Best Local Similarity			100.0%;	Pred. No. 9,2e-140;
Matches 273;			Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
1728 GTATCGGAGACAGAGAGAGCGGAGCAGAGCTCTTGGAGATTTCGCTTGGAAAGTCC			17627	

Db 21580 GTGATCGGAAACAGAGAGAGCGGAGCAGCTCTTGAGANTTTTCGTTTGAAAAGGCC 21639

Qy 1788 CAGTTCTTGTGCTACTTCACTAGCTCCAGAGGGCTGGATATTGAAAATGTGCACATG 1847

Db 21640 CAGTTCTTGTGCTACTTCACTAGCTCCAGAGGGCTGGATATTGAAAATGTGCACATG 21699

Qy 1848 TTATCAATTTTGATCTTCTCTCTACATGTAATGTTCATCGAATTGGGGCTACTG 1907

Db 21700 TTATCAATTTTGATCTTCTCTCTACATGTAATGTTCATCGAATTGGGGCTACTG 21759

Qy 1908 GTCTGTGGGAATACCTGCGAGAGCAATTTCTTTTGTATCTTGAATCGAATACCAT 1867

Db 21760 GTCTGTGGGAATACCTGCGAGAGCAATTTCTTTTGTATCTTGAATCGAATACCAT 21819

Qy 1968 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000

Db 21820 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 21852

RESULT 13

AC022265/c

LOCUS

DEFINITION

AC022265 204250 bp DNA linear HTG 13-AUG-2000

SEQUENCE, 16 unordered pieces.

AC022265

AC022265.3 GI:9719812

HTG; HTGS PHASE1; HTGS_DRAFT.

KEYWORDS

Source

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 204250)

Waterston,R.H.

TITLE

The sequence of Homo sapiens clone

JOURNAL

Unpublished

2 (bases 1 to 204250)

Waterston,R.H.

REFERENCE

Direct Submission

Submitted (27-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA.

COMMENT

On Aug 7, 2000 this sequence version replaced gi:7940395.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: MUGSC

Web site:http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H NH0332C17

----- Summary Statistics -----

Sequencing vector: M13; 87%

Sequencing vector: plasmid; 13%

Chemistry: Dye-Primer ET; 85% of reads

Chemistry: Dye-terminator Big Dye; 15% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 19649 bases at least Q40

Consensus quality: 200804 bases at least Q30

Consensus quality: 199226 bases at least Q30

Insert size: 205000; agarose-fp

Insert size: 202750; sum-of-ctrls

Quality coverage: 4.50 in Q20 bases; agarose-fp

Quality coverage: 4.65 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1062: contig of 1062 bp in length

FEATURES

source

1..1063

1162: gap of unknown length

2536: contig of 1374 bp in length

2636: gap of unknown length

5791: contig of 315 bp in length

5891: gap of unknown length

5892: contig of 5055 bp in length

10946: gap of unknown length

11047: gap of 4679 bp in length

15725: gap of unknown length

15826: gap of 5344 bp in length

21699: gap of unknown length

21700: gap of 6531 bp in length

27801: gap of unknown length

33276: gap of 5376 bp in length

33377: gap of unknown length

42711: contig of 9335 bp in length

42812: gap of unknown length

56046: contig of 13235 bp in length

56147: gap of unknown length

77263: contig of 21117 bp in length

77364: gap of unknown length

97974: contig of 20611 bp in length

98075: gap of unknown length

122272: contig of 24198 bp in length

122273: gap of unknown length

148199: contig of 25827 bp in length

148200: gap of unknown length

175035: contig of 26736 bp in length

175135: gap of unknown length

175136 204250: contig of 29115 bp in length.

Location/Qualifiers

1..204250

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="5"

/clone="RP11-332C17"

1..1062

/note="assembly_name:Contig12"

1163..2536

/note="assembly_name:Contig13"

2637..5791

/note="assembly_name:Contig14"

5892..10946

/note="assembly_name:Contig15"

11047..15725

/note="assembly_name:Contig16"

15826..21169

/note="assembly_name:Contig17"

21270..27800

/note="assembly_name:Contig18"

27901..33276

/note="assembly_name:Contig19"

33377..42711

/note="assembly_name:Contig20

clone_end:17

vector_side:right"

42812..56046

/note="assembly_name:Contig21"

56147..77263

/note="assembly_name:Contig22"

77364..97974

/note="assembly_name:Contig23"

98075..122272

/note="assembly_name:Contig24"

122373..148199

/note="assembly_name:Contig25"

148300..175035

/note="assembly_name:Contig26"

175136..204250

/note="assembly_name:Contig27"

ORIGIN

Query Match 7.8%; Score 174; DB 2; Length 204250;
 Best Local Similarity 100.0%; Pred. No. 1,3e-84;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1312 AAGATTGGTCTCAACAGATCAATCTAGTTTGGATGAAGTGCATGCATGTGGA 1371
 |||||
 Db 170192 AAGATTGGTCTCAACAGATCAATCTAGTTTGGATGAAGTGCATGCATGTGGA 170133
 |||||

Cy 1372 TATGGTTTGGTCCGAAATGAAGATTATTTCTTGGCCAGAAATCCATCAAGA 1431
 |||||
 Db 170132 TATGGTTTGGTCCGAAATGAAGATTATTTCTTGGCCAGAAATCCATCAAGA 170073
 |||||

Cy 1432 ACAGCGCCAAACCTTATGTCAGTCACTTTCCAGAGAAATCAAGGTT 1485
 |||||
 Db 170072 ACAGCGCCAAACCTTATGTCAGTCACTTTCCAGAGAAATCAAGGTT 170019
 |||||

RESULT 14
 AY100475 524 bp DNA linear MAM 15-JUN-2002
 LOCUS Equus caballus VASA-like protein gene, partial sequence.
 DEFINITION AY100475
 ACCESSION AY100475.1 GI:21429223
 VERSION
 KEYWORDS
 SOURCE Equus caballus (horse)
 ORGANISM Equus caballus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 REFERENCE 1 (bases 1 to 524)
 AUTHORS Woods,B.G., Ginther,O.J., Wentworth,A., Wentworth,B. and
 TITLE Equine VASA Homolog
 JOURNAL Unpublished
 2 (bases 1 to 524)
 AUTHORS Woods,B.G., Ginther,O.J., Wentworth,A., Wentworth,B. and
 Wiltbank,M.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2002) Animal Health and Biomedical Sciences,
 University of Wisconsin-Madison, 1656 Linden Drive, Madison, WI
 53706, USA
 FEATURES
 source
 1..524
 location/Qualifiers
 /organism="Equus caballus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9796"
 <1..>524
 /note="similar to Drosophila melanogaster VASA protein"

ORIGIN
 misc_feature
 /note="similar to Drosophila melanogaster VASA protein"

Query Match 2.6%; Score 58; DB 4; Length 524;
 Best Local Similarity 100.0%; Pred. No. 7.5e-20;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1997 GATGCTCAACAGATGTTCTCGCATGTTGGAAGAAATGCTTTAGTACATCAATTC 2054
 |||||
 Db 455 GATGCTCAACAGATGTTCTCGCATGTTGGAAGAAATGCTTTAGTACATCAATTC 512
 |||||

RESULT 15
 AC109144 211429 bp DNA linear HTG 22-FEB-2003
 LOCUS Mus musculus clone RP23-105P19, WORKING DRAFT SEQUENCE, 24
 DEFINITION Mus musculus clone RP23-105P19, WORKING DRAFT SEQUENCE, 24
 unoriented pieces.
 AC109144
 AC109144.3 GI:28467368
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 VERSION
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 211429)
 AUTHORS Birren,B., Nusbaum,C. and Lander,E.
 TITLE Mus musculus, clone RP23-105P19

JOURNAL
 2 (bases 1 to 211429)
 REFERENCES
 AUTHORS
 Unpublished
 Birren,B., Linton,J., Nusbaum,C., Lander,E., Allen,A., Allen,N.,
 Anderson,S., Barta,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,A., Chang,J., Chararo,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,J., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,N., Gage,D., Galagan,J., Gardyna,S.,
 Girde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kells,C., Lacroque,K., Lamas,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
 MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,
 McKean,P., McKernan,K., Meldrim,J., Meusels,L., Mihova,T.,
 Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norby,C.,
 Oldman,V., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Rector,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,C.,
 Roselli,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J.,
 Totham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 TITLE
 JOURNAL
 REFERENCES
 AUTHORS
 Direct Submission
 Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 211429)
 Anderson,S., Arachchi,H.M., Barta,N., Bastien,V., Bloom,T.,
 Boguslavsky,L., Boukhalter,B., Camarata,A., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagoopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
 MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meusels,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norby,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Rector,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Strubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Totham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 22, 2003 this sequence version replaced gi:22915576.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L16415
 Center clone name: 105 P 19
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye, 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 205419 bases at least Q40
 Consensus quality: 208054 bases at least Q30
 Consensus quality: 208754 bases at least Q20
 Insert size: 18200; agarose-efp
 Insert size: 209129; sum-of-contigs
 Quality coverage: 10.5 in Q20 bases; agarose-efp

Quality coverage: 9.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 20853: contig of 20853 bp in length
* 20854 20953: gap of 100 bp
* 20954 21633: contig of 680 bp in length
* 21634 21733: gap of 100 bp
* 21734 22390: contig of 657 bp in length
* 22391 22490: gap of 100 bp
* 22491 23404: contig of 914 bp in length
* 23405 23505: gap of 100 bp
* 23505 24146: contig of 641 bp in length
* 24146 24245: gap of 100 bp
* 24245 25058: contig of 813 bp in length
* 25059 25159: gap of 100 bp
* 25159 26138: contig of 980 bp in length
* 26139 26238: gap of 100 bp
* 26239 26898: contig of 660 bp in length
* 26899 26998: gap of 100 bp
* 26999 27867: contig of 868 bp in length
* 27867 27967: gap of 100 bp
* 27967 28591: contig of 624 bp in length
* 28591 28691: gap of 100 bp
* 28691 29543: contig of 853 bp in length
* 29544 29643: gap of 100 bp
* 29643 30707: contig of 1064 bp in length
* 30708 30807: gap of 100 bp
* 30808 32000: contig of 1193 bp in length
* 32001 32100: gap of 100 bp
* 32101 33156: contig of 1056 bp in length
* 33157 33256: gap of 100 bp
* 33257 34265: contig of 1009 bp in length
* 34266 34365: gap of 100 bp
* 34366 35371: contig of 1006 bp in length
* 35372 35471: gap of 100 bp
* 35472 36860: contig of 1389 bp in length
* 36861 36960: gap of 100 bp
* 36961 38773: contig of 1813 bp in length
* 38774 38873: gap of 100 bp
* 38874 40836: contig of 1963 bp in length
* 40837 40936: gap of 100 bp
* 40937 42700: contig of 1764 bp in length
* 42701 42800: gap of 100 bp
* 42801 59066: contig of 16266 bp in length
* 59067 59167: gap of 100 bp
* 59167 94920: contig of 35754 bp in length
* 94921 95020: gap of 100 bp
* 95021 151531: contig of 56511 bp in length
* 151532 151632: gap of 100 bp
* 151632 211429: contig of 59798 bp in length.
```

FEATURES

Source

```
1. .211429
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-105P19"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .20853
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
20954..21633
/note="assembly_fragment"
21734..22390
/note="assembly_fragment"
22491..23404
misc_feature
```

```
misc_feature /note="assembly_fragment"
23505..24145
/note="assembly_fragment"
misc_feature 24246..25058
/note="assembly_fragment"
misc_feature 25159..26138
/note="assembly_fragment"
misc_feature 26239..26898
/note="assembly_fragment"
misc_feature 26999..27866
/note="assembly_fragment"
misc_feature 27967..28590
/note="assembly_fragment"
misc_feature 28691..29543
/note="assembly_fragment"
misc_feature 29644..30707
/note="assembly_fragment"
misc_feature 30808..32000
/note="assembly_fragment"
misc_feature 32101..33156
/note="assembly_fragment"
misc_feature 33257..34265
/note="assembly_fragment"
misc_feature 34366..35371
/note="assembly_fragment"
misc_feature 35472..36860
/note="assembly_fragment"
misc_feature 36961..38773
/note="assembly_fragment"
misc_feature 38874..40836
/note="assembly_fragment"
misc_feature 40937..42700
/note="assembly_fragment"
misc_feature 42801..59066
/note="assembly_fragment"
misc_feature 59167..94920
/note="assembly_fragment"
misc_feature 95021..151531
/note="assembly_fragment"
misc_feature 151632..211429
/note="assembly_fragment"
```

Query Match 2.6%; Score 57; DB 2; Length 211429;
Best Local Similarity 100.0%; Pred. No. 1.9e-19; Mismatches 0; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 877 GACCTTTGAAGAGCTATCTCTGTGACACACTGATATACACACTTGGCTAAAGCTGG 933
DB 181836 GACCTTTGAAGAGCTATCTCTGTGACACACTGATATACACACTTGGCTAAAGCTGG 181892

Search completed: July 27, 2004, 16:33:06
Job time : 8608 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 10:49:52 ; Search time 879 Seconds
(without alignments)
10748.569 Million cell updates/sec

Title: US-09-714-865B-1
Perfect score: 2224
Sequence: 1 acttgagtcaccatg9999.....aagtcctggttgatgca 2224

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 212409041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: geneseq1980s.*
- 2: geneseq1990s.*
- 3: geneseq2000s.*
- 4: geneseq2001as.*
- 5: geneseq2001bs.*
- 6: geneseq2002s.*
- 7: geneseq2003as.*
- 8: geneseq2003bs.*
- 9: geneseq2003cs.*
- 10: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2224	100.0	2224	5	AAD06354 Human vas
2	1288	57.9	1984	9	ADB63498 Human CDN
3	579	26.0	1006	3	AAA44726 Human sec
4	393	17.7	677	4	AAS40979 CDNA enco
5	310	13.9	667	7	ABZ20094 Group III
6	278	12.5	591	4	AAS41534 CDNA enco
7	278	12.5	591	4	AAL01169 Human rep
8	278	12.5	591	4	ABL96630 Human tes
9	269	12.1	275	7	ABZ19496 Human tes
10	220	9.9	2194	4	AAL04793 Human rep
11	220	9.9	2194	4	ABL97687 Human tes
12	220	9.9	2195	4	AAL04794 Human rep
13	220	9.9	2195	4	ABL97688 Human tes
14	220	9.9	2197	4	AAL04795 Human rep
15	220	9.9	2197	4	ABL97689 Human tes
16	73	3.3	344	7	ABZ18691 Group III
17	60	2.7	60	6	ABN42617 Human spl
18	36	1.6	1383	7	ACA53957 Prokaryot
19	25	1.1	28	5	AAD06355 Human vas
20	23	1.0	2187	6	AB199243 Mouse isc
21	21	0.9	21	5	AAD06358 RT-PCR pr
22	21	0.9	1842	4	AAS53256 Haemophil
23	21	0.9	1842	7	ACA34056 Prokaryot

24	21	0.9	110000	2	AAT42063_02	Continuation (3 of
25	20	0.9	20	5	AAD06359	Aad06359 RT-PCR pr
26	20	0.9	20	5	AAD06356	Aad06356 PCR prime
27	20	0.9	2320	3	ACB37735	Aac37735 Arabidops
28	20	0.9	9832	6	ABL32657	Ab132657 Human Imm
29	20	0.9	96594	9	ADC85257	Adc85257 Mouse Pcp
30	20	0.9	96595	8	ADA02777	Ada02777 Mouse Pcp
31	20	0.9	96595	8	ADB72515	Adb72515 Mouse Pcp
32	20	0.9	110000	6	ABA03041_07	Continuation (8 of
33	19	0.9	28	5	AAD06357	Aad06357 PCR prime
34	19	0.9	138	4	ABA72946	Abat72946 Human toe
35	19	0.9	138	4	AA153370	AA153370 Probe #22
36	19	0.9	138	4	AAK47539	AAk47539 Human bon
37	19	0.9	138	4	AAK21377	AAk21377 Human bra
38	19	0.9	138	4	AB847275	Ab847275 Human liv
39	19	0.9	138	6	AB821638	Ab821638 Human gen
40	19	0.9	391	7	ACD92737	Accd92737 Human col
41	19	0.9	418	8	ACH48505	Act48505 Human leu
42	19	0.9	504	7	ADA70581	Ada70581 Rice gene
43	19	0.9	592	4	ABA60407	Ab60407 Human foe
44	19	0.9	592	4	AA140291	AA140291 Probe #89
45	19	0.9	592	4	AAK34573	AAk34573 Human bon

ALIGNMENTS

RESULT 1	
AAD06354	
ID	AAD06354 standard; CDNA; 2224 BP.
AC	AAD06354;
XX	
DT	10-AUG-2001 (first entry)
XX	
DE	Human vasa CDNA.
XX	
KW	Human; vasa; germ cell; gonad development; therapy; cancer; oral; brain;
KW	ovarian; biliary tract; breast; pancreas; prostate; colorectal; cervical;
KW	colon; lung; testis; renal; thyroid; oesophageal; endometrial; gastric;
KW	skin; liver; lymphoma; melanoma; glioblastoma; neuroblastoma; neoplasm;
KW	medulloblastoma; choriochorionoma; squamous cell carcinoma; leukemia;
KW	acute lymphocytic; myelogenous; multiple myeloma; Paget's disease;
KW	osteosarcoma; Acquired immune deficiency syndrome; AIDS; Bowen's disease;
KW	leiomyosarcoma; rhabdomyosarcoma; liposarcoma; Kaposi's sarcoma;
KW	fibrosarcoma; tumour; testicular; seminoma; ovarian; dysgerminoma;
KW	teratoma; mediastinal; intracranial; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Location/Qualifiers
FT	14..2188
FT	/*tag= a
FT	/product= "Human vasa protein"
FT	14..2185
FT	/*tag= b
FT	/note= "This region is specifically claimed as SEQ ID NO: 15 in claim 3 of the specification"
XX	
PN	WO200136445-A1.
XX	
PD	25-MAY-2001.
XX	
PF	16-NOV-2000; 2000WO-US031465.
XX	
PR	18-NOV-1999; 99US-0166394P.
XX	
PA	(BGMH) BRIGHAM & WOMENS HOSPITAL INC.
XX	
PI	Castillion DH;
XX	
DR	WPI, 2001-355606/37.
DR	P-PSDB; AAE02417.

XX Novel vasa polynucleotides useful in the diagnosis or treatment of
PT conditions characterized by aberrant expression and/or presence of mutant
PT forms of vasa polynucleotides or polypeptides.
XX
PS Claim 1; Page 52-53; 66pp; English.
XX
CC The present sequence is human vasa cDNA that has germ cell specific
CC expression and is believed to play a determinative role in gonad
CC development. Germ cells are specialised to produce haploid gametes in
CC multicellular organisms. Vasa is useful in the diagnosis or treatment of
CC conditions characterised by its aberrant expression and/or the presence
CC of its mutant forms. The conditions include cancers such as biliary
CC tract, brain, breast, colon, ovarian, pancreas, prostate, colorectal,
CC oral, liver, lung, skin, basocellular, testis, renal, thyroid, cervical,
CC endometrial, oesophagel and gastric, lymphomas, melanomas,
CC glioblastomas, neuroblastomas, medulloblastomas, chorioarctinoma,
CC squamous cell carcinoma, haematological neoplasms, acute lymphocytic and
CC myelogenous leukaemia, multiple myeloma, Acquired immune deficiency
CC syndrome (AIDS) associated leukemias, intraepithelial neoplasms, Bowen's
CC disease, Paget's disease, sarcomas such as leiomyosarcoma,
CC rhabdomyosarcoma, liposarcoma, fibrosarcoma, Kaposi's sarcoma and
CC osteosarcoma, tumours such as testicular tumour (eg. seminoma), ovarian
CC tumour (eg. dysgerminoma or teratoma) and tumour of an extragonadal
CC tissue (eg. mediastinal or an intracranial tumour)
XX
SQ Sequence 2224 BP; 678 A; 385 C; 550 G; 611 T; 0 U; 0 Other;
Query Match 100.0%; Score 2224; DB 5; Length 2224;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 601 TACTTCCAAAGCAAGAGTGGCAAGTGAACGAGGTGGTTACAAAGTTTAAATGA 660
QY AGAAGTAATTAACAGAGCTCTGGAAAAGAAATCTTGGAACTCAAGACAGAGAGAGAG 720
DB 661 AGAAGTAATTAACAGAGCTCTGGAAAAGAAATCTTGGAACTCAAGACAGAGAGAGAG 720
QY 721 TAGTGATATCTAAGACCAAGAAAGTGAACCTTCAACCCCTCCACCTGAGAGTGA 780
QY 781 CTCATCTTTGACATTAATGACAGAGCATTAATCTTGAACAAATACGACATATTTCTGT 840
DB 781 CTCATCTTTGACATTAATGACAGAGCATTAATCTTGAACAAATACGACATATTTCTGT 840
QY 841 GGAAGTGTGGAATATATGACACAGAGAAATTCGATCTTTGANAAGCTAATCTCTG 900
DB 841 GGAAGTGTGGAATATATGACACAGAGAAATTCGATCTTTGANAAGCTAATCTCTG 900
QY 901 TCACACACTGATTAACAAATGCTTAAGCTGTTACTAATCTTCTGTCGAAA 960
DB 901 TCACACACTGATTAACAAATGCTTAAGCTGTTACTAATCTTCTGTCGAAA 960
QY 961 ATACAGATTTCTTATCTACTTGCAGAGACAGATTTGATGCTTGTGCTCAACAGGCTC 1020
DB 961 ATACAGATTTCTTATCTACTTGCAGAGACAGATTTGATGCTTGTGCTCAACAGGCTC 1020
QY 1021 TGGGAAGACTGCGCTTTCTCTACCAATTTTGGCTCATATGATGATGATGATGATGATG 1080
DB 1021 TGGGAAGACTGCGCTTTCTCTACCAATTTTGGCTCATATGATGATGATGATGATGATG 1080
QY 1081 TGCCAGTGTGTTTAAAGATTGACAGAAACAGAGTATTTATGTCACCACTGACGA 1140
DB 1081 TGCCAGTGTGTTTAAAGATTGACAGAAACAGAGTATTTATGTCACCACTGACGA 1140
QY 1141 ATTGTCACACAGATTTATTTGGAAGCCAGAAAATTTCTTTGGGACTGTGTAAGAGC 1200
DB 1141 ATTGTCACACAGATTTATTTGGAAGCCAGAAAATTTCTTTGGGACTGTGTAAGAGC 1200
QY 1201 TGTGTTATATATATGAGGGAACCCAGCTGGACATTCATTCGACAAATAGTACAGGCTG 1260
DB 1201 TGTGTTATATATATGAGGGAACCCAGCTGGACATTCATTCGACAAATAGTACAGGCTG 1260
QY 1261 TAAATATATATATGCTCTCTCTGGAAGACATGATGATGATGATGATGATGATGATG 1320
DB 1261 TAAATATATATATGCTCTCTCTGGAAGACATGATGATGATGATGATGATGATGATG 1320
QY 1321 TCTCAACACAGATCAAAATATAGTTTGAATGAAGCTGATGATGATGATGATGATG 1380
DB 1321 TCTCAACACAGATCAAAATATAGTTTGAATGAAGCTGATGATGATGATGATGATGATG 1380
QY 1381 TGTGTCAGAAATGAAGAAATTAATTTCTTGGCCAGAAATCCATCAAGAGACAGGCGCA 1440
DB 1381 TGTGTCAGAAATGAAGAAATTAATTTCTTGGCCAGAAATCCATCAAGAGACAGGCGCA 1440
QY 1441 AACCCTTATGTCAGTCAACTTTCCAGAGAAATCAAGAGTGGCTCAGAGTTT 1500
DB 1441 AACCCTTATGTCAGTCAACTTTCCAGAGAAATCAAGAGTGGCTCAGAGTTT 1500
QY 1501 AAAGTCAAAATATCTGTTTGTGCTGTGGAACAAGTGGGAGCAATGAGAGATGTCA 1560
DB 1501 AAAGTCAAAATATCTGTTTGTGCTGTGGAACAAGTGGGAGCAATGAGAGATGTCA 1560
QY 1561 GCAGACGCTTCCAAAGTGGCCAGTCTCAAAAAGAGAAAGTGTGTAATTTCTGCG 1620
DB 1561 GCAGACGCTTCCAAAGTGGCCAGTCTCAAAAAGAGAAAGTGTGTAATTTCTGCG 1620
QY 1621 AAACATAGGGGATGAAGAACTATGCTTTGTTGAACATTAAGAAAAAGCAGATTTTAC 1680
DB 1621 AAACATAGGGGATGAAGAACTATGCTTTGTTGAACATTAAGAAAAAGCAGATTTTAC 1680
QY 1661 TGCAACTTTTCTTGTCAAGAAAAATATCACTCAAGATTCATGATGATGATGATGATG 1740
DB 1661 TGCAACTTTTCTTGTCAAGAAAAATATCACTCAAGATTCATGATGATGATGATGATG 1740

Db	1661	TGCAACITTTCTTTGTGTCAGAAAATAATATCACTCAAGTATCCATGTCATCGGGAACA	1740
QY	1741	GAGAGAGGGGGAGCAAGCTCTTGAGATTTTCGCTTGGAAAGGCCCAAGTCTTGTTGC	1800
Db	1741	CAGAGAGCGGGAGCAAGCTCTTGAGATTTTCGCTTGGAAAGGCCCAAGTCTTGTTGC	1800
QY	1801	TACTTCAGTAGCTGCCAGAGGGCTGGATATTGAAATGTGCACATGTTATCAATTTTGA	1860
Db	1801	TACTTCAGTAGCTGCCAGAGGGCTGGATATTGAAATGTGCACATGTTATCAATTTTGA	1860
QY	1861	TCTTCCTCTCCACATGATGAAATAGTTCATCGAATTTGGGGTACTGGTCTGTGGGAA	1920
Db	1861	TCTTCCTCTCCACATGATGAAATAGTTCATCGAATTTGGGGTACTGGTCTGTGGGAA	1920
QY	1921	TACTGGCAGACCAATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCACAGCCTCT	1980
Db	1921	TACTGGCAGACCAATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCACAGCCTCT	1980
QY	1981	AGTAAAGATATTGACAGATGCTGCAACAGAGATGCTCCGATGTTGGAAGAAATGGCTT	2040
Db	1981	AGTAAAGATATTGACAGATGCTGCAACAGAGATGCTCCGATGTTGGAAGAAATGGCTT	2040
QY	2041	TAGTACATACATTCCTGGCTTCAGTGTAGTACACAGAGAAACGTGTTGCATCAGTTGA	2100
Db	2041	TAGTACATACATTCCTGGCTTCAGTGTAGTACACAGAGAAACGTGTTGCATCAGTTGA	2100
QY	2101	TACCAAGAAAGGGCAAGAGCACTTTGGAACACAGCTGGGTTTCTTCTTACAGCTCCCAA	2160
Db	2101	TACCAAGAAAGGGCAAGAGCACTTTGGAACACAGCTGGGTTTCTTCTTACAGCTCCCAA	2160
QY	2161	TCCAGTAGATANTAGTCATGGAGTTAAAGCCAAACATCCTTCAAGCTGTGGTTTGA	2220
Db	2161	TCCAGTAGATANTAGTCATGGAGTTAAAGCCAAACATCCTTCAAGCTGTGGTTTGA	2220
QY	2221	TGCA 2224	
Db	2221	TGCA 2224	

XX	RESULT 2
XX	ADBB63498
ID	ADBB63498 standard; cDNA; 1984 BP.
AC	
XX	ADBB63498;
DT	04-DEC-2003 (first entry)
DE	Human cDNA encoding clone TESTI20227380.
XX	
KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy; tissue regeneration; cell regeneration; membrane protein; signal transduction-related protein; transcription-related osteoporosis; neurological disease; cancer; tumour.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
CDS	20..1747
FT	/tag= a
FT	/product= "Clone TESTI20227380 protein"
XX	
PN	EPI308459-A2.
XX	
PD	07-MAY-2003.
XX	
PF	28-MAR-2002; 2002EP-00007401.
XX	
PR	05-NOV-2001; 2001UP-00378298.
XX	
PA	25-JAN-2002; 2002US-00350978.
XX	
PA	(HELI-) HELIX RES INST.
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.
XX	

Sequence 1984 BP, 612 A, 357 C, 448 G, 567 T, 0 U, 0 Other;

Query Match	57.9%	Score 1288;	DB 9;	Length 1984;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 1488;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0
QY	733	AGACCAAAAGTACCTACATACCCCTCCTCCACTGAGATGAGACTCCATCTTTC	792	
Db	292	AGACCAAAAGTACCTACATACCCCTCCTCCACTGAGATGAGACTCCATCTTTC	351	
QY	793	ACATTATCAGACAGCATTAACCTTCGCAAAATGACACATATCTTGAGAGTGTCTG	852	
Db	352	ACATTATCAGACAGCATTAACCTTCGCAAAATGACACATATCTTGAGAGTGTCTG	411	
QY	853	ACATGATGACCACCCAGCAATTCGACTTTGAAAGACTAATCTGTGACACACTGA	912	
Db	412	ACATGATGACCACCCAGCAATTCGACTTTGAAAGACTAATCTGTGACACACTGA	471	
QY	913	TACCAACAATGTGAAAGCTGTATACATAGCTTACCTCGTGCACAAATATACAGATTC	972	
Db	472	TACCAACAATGTGAAAGCTGTATACATAGCTTACCTCGTGCACAAATATACAGATTC	531	
QY	973	TATCATACTTGCAGACGAGATTTGATGAGCTTGTCTCAACAGGGTCTGGAGAGCTGC	1033	
Db	532	TATCATACTTGCAGACGAGATTTGATGAGCTTGTCTCAACAGGGTCTGGAGAGCTGC	591	
QY	1033	GACCTTTCTCCATCAAAATTTGGCTCATATGATGACATGATGAAATACGCCAGTCGTT	1097	
Db	592	GACCTTTCTCCATCAAAATTTGGCTCATATGATGACATGATGAAATACGCCAGTCGTT	651	
QY	1093	TAAAGAGTTGCAGAGAACGAGTGTATATTTGAGACCAACAACCTCGAATTTGGTCAACCA	1155	
Db	652	TAAAGAGTTGCAGAGAACGAGTGTATATTTGAGACCAACAACCTCGAATTTGGTCAACCA	711	
QY	1153	GATTATTTTGAAGCGAAATTTCTTTTGGACTGTGTGAGACTGTGTGATATATA	1211	

Db 712 GATTATTTGGAGACCCAGAAAATTTCTTTGGGACTGTGTAGAGCTGTGTATATA 771

Qy 1213 TGGGGAGACCCAGCTGGGACATTCATTCGACCAATAGTACAGGCTGTATATATATG 1272

Db 772 TGGGGAGACCCAGCTGGGACATTCATTCGACCAATAGTACAGGCTGTATATATATG 831

Qy 1273 TGTCTCTCTGGAGAGCTGATGATATCATAGGCAAGAAAAGATTTGGTCTCAACAGAT 1332

Db 832 TGTCTCTCTGGAGAGCTGATGATATCATAGGCAAGAAAAGATTTGGTCTCAACAGAT 891

Qy 1333 CAATCTCTATGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 1392

Db 892 CAATCTCTATGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 951

Qy 1393 GAAGAAATTAATTTCTTGGCCAGAGATGCCATCAAGAAACAGCCCAACCTTATGTT 1452

Db 952 GAAGAAATTAATTTCTTGGCCAGAGATGCCATCAAGAAACAGCCCAACCTTATGTT 1011

Qy 1453 CAGTGGACCTTTCCAGAGAGAAATCAAGAGTGGCTGAGAGATTTTAAAGTCAATTA 1512

Db 1012 CAGTGGACCTTTCCAGAGAGAAATCAAGAGTGGCTGAGAGATTTTAAAGTCAATTA 1071

Qy 1513 TCTGTTGTGCTGTGTGACCAAGTGGGTGAGCATGTAGAGATTTTCAAGACCGTTCT 1572

Db 1072 TCTGTTGTGCTGTGTGACCAAGTGGGTGAGCATGTAGAGATTTTCAAGACCGTTCT 1131

Qy 1573 CCAAGTTGGCCAGTCTTCCAAAAGAGAAAAGCTGCTGAAATTTGCGGAAACATATGGGGA 1632

Db 1132 CCAAGTTGGCCAGTCTTCCAAAAGAGAAAAGCTGCTGAAATTTGCGGAAACATATGGGGA 1191

Qy 1633 TGAAGAACTATGCTTGTGTGAAAATAAGAAAAGAGATTTTACTGCAACTTTCT 1692

Db 1192 TGAAGAACTATGCTTGTGTGAAAATAAGAAAAGAGATTTTACTGCAACTTTCT 1251

Qy 1693 TTGTCAAGAAAATATCAACTCAAGTATCCATGTGATCGGGAAACAGAGAGAGGGGA 1752

Db 1252 TTGTCAAGAAAATATCAACTCAAGTATCCATGTGATCGGGAAACAGAGAGAGGGGA 1311

Qy 1753 GCAAGCTCTTGGAGATTTGCTTGGAAAGTCCAGATTTCTGTGCTACTTCACTAGTAC 1812

Db 1312 GCAAGCTCTTGGAGATTTGCTTGGAAAGTCCAGATTTCTGTGCTACTTCACTAGTAC 1371

Qy 1813 TGCAGAGGGCTGATATGAAAATGTGCAACATGTTATCAATTTGATCTTCTTCTAC 1872

Db 1372 TGCAGAGGGCTGATATGAAAATGTGCAACATGTTATCAATTTGATCTTCTTCTAC 1431

Qy 1873 CATTTGATGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1932

Db 1432 CATTTGATGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491

Qy 1933 AATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCAACAGCTCTAGTAAAGTATT 1992

Db 1492 AATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCAACAGCTCTAGTAAAGTATT 1551

Qy 1993 GACAGATGCTCAACAGATGTTCTGATGATGATGATGATGATGATGATGATGATGAT 2052

Db 1552 GACAGATGCTCAACAGATGTTCTGATGATGATGATGATGATGATGATGATGATGAT 1611

Qy 2053 TCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2112

Db 1612 TCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1671

Qy 2113 CAAGAGCACTTTGAACACAGCTGGGTTTCTTCTTCAAGAGCTTCCCAATCCAGTAATGA 2172

Db 1672 CAAGAGCACTTTGAACACAGCTGGGTTTCTTCTTCAAGAGCTTCCCAATCCAGTAATGA 1731

Qy 2173 TGAATCATGGGATTAAGCCAAAATCTTCAAGTCTGTGTTTGAATGA 2224

Db 1732 TGAATCATGGGATTAAGCCAAAATCTTCAAGTCTGTGTTTGAATGA 1783

ID AAA44726 standard; cDNA; 1006 BP.

AC AAA44726;

XX 21-AUG-2000 (first entry)

DT Human secreted expressed sequence tag SEQ ID NO:1301.

DE Human; mouse; chicken; rat; secreted expressed sequence tag; SEST; expressed sequence tag; SEST; probe; chemotactic; proliferative; immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic; chemokine; anti-inflammatory; cytoskeletal; antibacterial; antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian; antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

OS WO200021991-A1.

PN 20-APR-2000.

XX 15-OCT-1999; 99WO-US024206.

XX 15-OCT-1998; 98US-010436P.

PR (GENY) GENETICS INST INC.

PA Jacobe K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C, Meiberg D, Treacy M, Bowman MR;

XX WPI; 2000-317938/27.

DR Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders.

PT Claim 1; Page 540; 803pp; English.

XX AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic; thrombolytic; anti-inflammatory; cytoskeletal; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor; osteoprotective; neuroprotective; nootropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention

CC Sequence 1006 BP; 314 A; 181 C; 251 G; 260 T; 0 U; 0 Other;

XX

Query Match 26.0%; Score 579; DB 3; Length 1006;

Best Local Similarity 99.8%; Pred. No. 5.3e-276;

Wed Jul 28 11:42:29 2004

us-09-714-865b-1.rng

Page 5

Matches 629; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Qy	510	ATAATGACTTAGCCAGACGATGTAATGACGACGCGGCGCTTTTGGTCTCTGAA	569	
Db	377	ATAATGACTTAGCCAGACGATGTAATGACGACGCGCTTTTGGTCTCTGAA	436	
Qy	570	GACCAATTAATGAGGACAGGTAATGTAATCTTCTCAAGACGAAAGTGGCAGTGA	629	
Db	437	GACCAATTAATGAGGACAGGTAATGTAATCTTCTCAAGACGAAAGTGGCAGTGA	496	
Qy	630	GTGAACGAGGTGTGTAACAAAGTTTAAATGAAGAATTAATACGCTCTGGAAAGATT	689	
Db	497	GTGAACGAGGTGTGTAACAAAGTTTAAATGAAGAATTAATACGCTCTGGAAAGATT	556	
Qy	690	CTTGAAGTCAGAAAGCAGAAAGAGAGAAAGTACTGTAATCTCAAGACCAAAAGTACT	749	
Db	557	CTTGAAGTCAGAAAGCAGAAAGAGAGAAAGTACTGTAATCTCAAGACCAAAAGTACT	616	
Qy	750	ACATACCCCTCCCTCCACCTGAGATGAGACTCCATCTTTGCACATTATCAACAGGCA	809	
Db	617	ACATACCCCTCCCTCCACCTGAGATGAGACTCCATCTTTGCACATTATCAACAGGCA	676	
Qy	810	TAACTTGCACAAATACGACACTATCTTGTGAAAGTGTCTGACATGATGACACACAG	869	
Db	677	TAACTTGCACAAATACGACACTATCTTGTGAAAGTGTCTGACATGATGACACACAG	736	
Qy	870	CAATCTGACTTTTGAAGAAGCTATCTCTGTGACACTGATGATACCAATGCTTAAG	929	
Db	737	CAATCTGACTTTTGAAGAAGCTATCTCTGTGACACTGATGATACCAATGCTTAAG	796	
Qy	930	CTGGTTACTTAAGCTTACTCCTGTGCAAAATCTCAGTATCTCAGTATCTTGGAGAC	989	
Db	797	CTGGTTACTTAAGCTTACTCCTGTGCAAAATCTCAGTATCTCAGTATCTTGGAGAC	856	
Qy	990	GAGATTGATGCTTGTGCTTGAACAGAGGTCTGGGAAGCTGGGCTTTTCTCTACCA	1049	
Db	857	GAGATTGATGCTTGTGCTTGAACAGAGGTCTGGGAAGCTGGGCTTTTCTCTACCA	916	
Qy	1050	TTTTGGCTCATATGATGATGATGATGATTAATCTGCACTGCTTTTAAAGTTGACGAA	1109	
Db	917	TTTTGGCTCATATGATGATGATGATGATTAATCTGCACTGCTTTTAAAGTTGACGAA	976	
Qy	1110	CAGAGTATATTAATGTAAGCACTGAG	1139	
Db	977	CAGAGTATATTAATGTAAGCACTGAG	1006	

PR	04-FEB-2000;	2000US-0180628P.
PR	24-FEB-2000;	2000US-0184664P.
PR	02-MAR-2000;	2000US-0186350P.
PR	16-MAR-2000;	2000US-0189874P.
PR	17-MAR-2000;	2000US-0190076P.
PR	18-APR-2000;	2000US-0198123P.
PR	19-MAY-2000;	2000US-0205155P.
PR	07-JUN-2000;	2000US-0209467P.
PR	28-JUN-2000;	2000US-0214886P.
PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225477P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226686P.
PR	23-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0233063P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0234984P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235835P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236803P.
PR	02-OCT-2000;	2000US-0236803P.

RESULT 4
AAS40979
ID AAS40979 standard; cDNA; 677 BP.
XX
AC AAS40979;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #195.
XX
KM Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KM ligase; hyperproliferative disorder; immunodeficiency disorder;
KM autoimmune disorder; neurological disorder; metabolic disorder;
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;
KM blood-related disorder; infectious disorder; gene therapy; cytostatic;
KM anti arthritic; nephrotropic; anticoagulant; ss.
XX
XX Homo sapiens.
XX OS
XX PN WO200155301-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001239.
XX PR 31-JAN-2000; 2000US-0179065P

KM recombinant DNA technology; gene; ss.
XX
XX Homo sapiens.
OS
XX WO200278516-A2.
PN
XX 10-OCT-2002.
PD
XX 28-MAR-2002; 2002WO-US010421.
PF
XX 30-MAR-2001; 2001US-0280255P.
PR 28-AUG-2001; 2001US-0315563P.
PR 09-JAN-2002; 2002US-0347313P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Wang T, Wang S, Bangur CS, Gaiger A;
PI
XX WPI; 2003-058387/05.
DR
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
PT preventing and treating cancer expressing CT or CP mRNA antigens, and, in
PT virology, immunology, microbiology, molecular biology and recombinant DNA
PT techniques.
XX
XX Claim 1; SEQ ID NO 2520; 207pp; English.
PS
XX
XX AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and
CC AB054446 to AB054472 represent protein (II) sequences, from the present
CC invention. (I) and (II) have cytostatic activity and can be used in gene
CC therapy and vaccines. (I), (II), antibodies and compositions from the
CC present invention are useful for diagnosing, preventing and treating
CC cancer, which expresses CT or CP mRNA antigens. They are useful for
CC stimulating immune response. They can also be useful in virology,
CC immunology, microbiology, molecular biology and recombinant DNA
CC techniques. N.B. The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pcl_sequences
XX
SQ Sequence 667 BP; 215 A; 146 C; 155 G; 150 T; 0 U; 1 Other;
Query Match 13.9%; Score 310; DB 7; Length 667;
Best Local Similarity 100.0%; Pred. No. 1e-142;
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 655 AATGAAGAAGTAATACAGGCTCTGGAAGAATTCTTGAAGTCAGAGCAGAGG 714
DB 1 AATGAAGAAGTAATACAGGCTCTGGAAGAATTCTTGAAGTCAGAGCAGAGG 60
QY 715 AGAAGTAGTATCTCAAGACCAAGAGTACTTACATACCCCTCTCTCACTTGA 714
DB 61 AGAAGTAGTATCTCAAGACCAAGAGTACTTACATACCCCTCTCTCACTTGA 120
QY 775 TGAGACTCCATCTTTGCACTTTTCAGACAGGATTAACCTTCAGAAATACGACTAT 834
DB 121 TGAGACTCCATCTTTGCACTTTTCAGACAGGATTAACCTTCAGAAATACGACTAT 180
QY 835 TCTTGTGGAAGTGTCTGACATGATGACACCAAGCAATTCGTGACTTTGAAGAAGCTAA 894
DB 181 TCTTGTGGAAGTGTCTGACATGATGACACCAAGCAATTCGTGACTTTGAAGAAGCTAA 240
QY 895 TCTTGTGACACATGATTAACCAATTCGTAAGGCTTAAAGCTTACTCTCTGT 954
DB 241 TCTTGTGACACATGATTAACCAATTCGTAAGGCTTAAAGCTTACTCTCTGT 300
QY 955 GCAAAATATAC 964
DB 301 GCAAAATATAC 310

RESULT 6
AAS41534
ID AAS41534 standard; cDNA; 591 BP.

XX
AC AAS41534;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #750.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; ss.
OS
XX Homo sapiens.
XX
XX WO200155301-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001239.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-MAR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205125P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216809P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225478P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227099P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.

PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 03-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-465566/50.
 DR P-PSDB; AAU23664.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing, preventing,
 PT treating neural, immune system, muscular, reproductive, pulmonary,
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.
 XX
 PS Claim 4; SEQ ID NO 760; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23614), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
 CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
 CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
 CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
 CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
 CC infectious disorders (e.g. influenza). The polynucleotides of the
 CC invention can also be used in gene therapy. AA54785-AA54184 represent
 CC cDNA sequences encoding for the novel human enzyme polypeptides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
 XX
 SQ Sequence 591 BP; 203 A; 100 C; 111 G; 176 T; 0 U; 1 Other;
 Query Match 12.5%; Score 278; DB 4; Length 591;
 Best Local Similarity 99.7%; Pred. No. 7.2e-127;
 Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1896 TTGGGCGTACCTGCTGCTGGGGAATCTGCGAGAGCAATTCCTTTTGGATCTGAAT 1955
 DB 30 TTGGGCGTACCTGCTGCTGGGGAATCTGCGAGAGCAATTCCTTTTGGATCTGAAT 89
 QY 1956 CGGATAACCATTTAGACAGAGCTCTAGTAAGATTTAGACAGATGCTCAACAGATGTC 2015
 DB 90 CGGATAACCATTTAGACAGAGCTCTAGTAAGATTTAGACAGATGCTCAACAGATGTC 149
 QY 2016 CTGATGTTGGAAGAAATGCTTGTGATACATATCTGCTCAGGTAGTACAA 2075
 DB 150 CTGATGTTGGAAGAAATGCTTGTGATACATATCTGCTCAGGTAGTACAA 209
 QY 2076 GAGGAAACGTTTGGATCATGATACAGAAAGGCAAGACACTTTGAACAGACTG 2135
 DB 210 GAGGAAACGTTTGGATCATGATACAGAAAGGCAAGACACTTTGAACAGACTG 269
 QY 2136 GGTTCCTTCTTCAACGAGCTCCCAATCATAGATGATGATGATGATTAAGCCAAA 2195
 DB 270 GGTTCCTTCTTCAACGAGCTCCCAATCATAGATGATGATGATGATTAAGCCAAA 329
 QY 2196 ACATCCTCAAGTCTGCTTGGATGCA 2224

Db 330 ACATCCTTCAGTCTGCTTTGATGCA 358

RESULT 7
AAL01169
ID AAL01169 standard; cDNA; 591 BP.
XX
AC AAL01169;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SHQ ID NO: 1170.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN MO200155320-A2.
XX
PD 02-AUG-2001.
XX
FF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225271P.
PR 14-AUG-2000; 2000US-0225474P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229333P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 06-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0244675P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251719P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX P-PSDB; AAM95199.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; SEQ ID NO 1170; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention
XX
XX Sequence 591 BP; 203 A; 100 C; 111 G; 176 T; 0 U; 1 Other;
SQ
Query Match 12.5%; Score 278; DB 4; Length 591;
Best Local Similarity 99.7%; Pred. No. 7, 2e-127;
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1896 TTGGGCGTACTGCTGTTGGGAAATCTGGCAGACCAATTCCTTTTGTGATCTTGAAT 1955
DB 30 TTGGGCGTACTGCTGTTGGGAAATCTGGCAGACCAATTCCTTTTGTGATCTTGAAT 89
QY 1956 CGGATACCATTTAGACAGACCCCTCTAGTAAAGTATTGACAGTCTCAACAGATGTTTC 2015
DB 90 CGGATACCATTTAGACAGACCCCTCTAGTAAAGTATTGACAGTCTCAACAGATGTTTC 149
QY 2016 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCACTGTTAGTACAA 2075
DB 150 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCACTGTTAGTACAA 209
QY 2076 GAGGAAACGTTGTCATGTCATGATACAGGAAAGGCAAGAGCACTTTGAACAGAGCTG 2135
DB 210 GAGGAAACGTTGTCATGTCATGATACAGGAAAGGCAAGAGCACTTTGAACAGAGCTG 269
QY 2136 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGATGATGATGATGATGATGATGATGAT 2195
DB 270 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGATGATGATGATGATGATGATGATGAT 329
QY 2196 ACACTCTTCAAGTCTGTGTTGATGCA 2224
DB 330 ACACTCTTCAAGTCTGTGTTGATGCA 358
RESULT 8
ABL96630
ID ABL96630 standard; cDNA; 591 BP.
XX
XX ABL96630;
XX
XX 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 298.

XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX W020015317-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001329.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198133P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217467P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228242P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.

PR 14-SEP-2000; 2000US-0231065P.
PR 21-SEP-2000; 2000US-0234232P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241222P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 21-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
DR
PT Nucleic acids encoding 973 human testicular antigen polypeptides, useful
XX for preventing, diagnosing and/or treating testicular cancer.
XX
PS Claim 1; SEQ ID NO 298; 766bp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
XX invention
XX
SQ Sequence 591 BP; 203 A; 100 C; 111 G; 176 T; 0 U; 1 Other;
Query Match 12.5%; Score 278; DB 4; Length 591;
Best Local Similarity 99.7%; Pred. No. 7.2e-127;
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1896 TTGGGCGTACTGATGCTGTGGGAATCTGGGAGAGCAATTCCTTTTGTATCTTGAT 1955
Db 30 TTGGGCGTACTGATGCTGTGGGAATCTGGGAGAGCAATTCCTTTTGTATCTTGAT 89
QY 1956 CGGATTAACCATTTAGACACAGCCTCTAGTAAAGATTTGACAGATGCTCAACAGATGCTC 2015
Db 90 CGGATTAACCATTTAGACACAGCCTCTAGTAAAGATTTGACAGATGCTCAACAGATGCTC 149
QY 2016 CTGCATGCTTGAAGAAATTCCTTAGACATATCCCTGGCTTCAGTGTATACAA 2075
Db 150 CTGCATGCTTGAAGAAATTCCTTAGACATATCCCTGGCTTCAGTGTATACAA 209
QY 2076 GAGGAAGCGTTTGCATCAGTTGATATACAGAAAGGAGCACTTTGAAACAGCTG 2135
Db 210 GAGGAAGCGTTTGCATCAGTTGATATACAGAAAGGAGCACTTTGAAACAGCTG 269
QY 2136 GGTTCCTTCCTTCACAGAGTCCCAATCCAGTATGATGATGATGGATTTAAAGCCAA 2195
Db 270 GGTTCCTTCCTTCACAGAGTCCCAATCCAGTATGATGATGATGGATTTAAAGCCAA 329
QY 2196 ACATCCTTCAGTCTGTGTTTGTATGCA 2224
Db 330 ACATCCTTCAGTCTGTGTTTGTATGCA 358
RESULT 9
ABZ19496
ID ABZ19496 standard; cDNA; 275 BP.
XX
AC ABZ19496;
XX
DT 23-JAN-2003 (first entry)
XX
DE Group III cDNA cancer related clone SEQ ID NO:1922.
XX
KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.
OS Homo sapiens.
XX
XX
XX WO200278516-A2.

PD 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US010421.
XX
XX 30-MAR-2001; 2001US-0280255P.
XX 28-AUG-2001; 2001US-0315563P.
XX 09-JAN-2002; 2002US-0347313P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang S, Bangur CS, Gaiger A,
XX WPI; 2003-058387/05.
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and in
XX virology, immunology, microbiology, molecular biology and recombinant DNA
XX techniques.
XX
XX Claim 1; SEQ ID NO 1922; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques. N.B. The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at http://wipo.int/pub/published_pct_sequences
XX
XX Sequence 275 BP; 87 A; 46 C; 64 G; 78 T; 0 U; 0 Other;
SQ
Query Match 12.1%; Score 269; DB 7; Length 275;
Best Local Similarity 100.0%; Pred. No. 2.1e-122;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 CATTCAATTGCAATAGTACAAAGCGTGTAAATATTTATGCTCTCTCTGGAAGACTG 1291
Db 7 CATTCAATTGCAATAGTACAAAGCGTGTAAATATTTATGCTCTCTCTGGAAGACTG 66
QY 1292 ATGGATTCATAGGCAAGAAAGATGGTCTCAACAGATCAAACTAGTTTGGAT 1351
Db 67 ATGGATTCATAGGCAAGAAAGATGGTCTCAACAGATCAAACTAGTTTGGAT 126
QY 1352 GAAGCTGATGCATGTTGGATATGGGTTTGGTCCAGAAATGAAAGATTATTTCTTGC 1411
Db 127 GAAGCTGATGCATGTTGGATATGGGTTTGGTCCAGAAATGAAAGATTATTTCTTGC 186
QY 1412 CCGAGATGCGCATCAAGGAACGCGCAACCTATGTTCAAGTCACTTTCCAGAG 1471
Db 187 CCGAGATGCGCATCAAGGAACGCGCAACCTATGTTCAAGTCACTTTCCAGAG 246
QY 1472 GAATTCAAAGGTTGCTGCAGAGTTT 1500
Db 247 GAATTCAAAGGTTGCTGCAGAGTTT 275
RESULT 10
AAL04793
ID AAL04793 standard; DNA; 2194 BP.
XX
XX AAL04793;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 7481.
XX
XX Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX

OS Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198133P.
XX 19-MAY-2000; 2000US-0205152P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.

[illegible]

XX	PI	Rosen CA, Barash SC, Ruben SM;
XX	DR	WPI; 2001-46570/50.
XX	PT	Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.
XX	PS	Disclosure; SEQ ID NO 7481; 1297bp + Sequence Listing; English.
XX	CC	The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention
XX	CC	
XX	CC	
SQ	Sequence	2194 BP; 624 A; 349 C; 400 G; 821 T; 0 U; 0 Other;
Query Match		9.3%; Score 220; DB 4; Length 2194;
Best Local Similarity		100.0%; Pred. No. 3,9e-98;
Matches 220;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1781	AAGTGGCCAGTTCTTGTTGCTACTCAATGCTGCCAGAGGGCTGATATTGAATGTG 1840
Db	1	AAGGGCCAGATTGTTGTTGCTACTTCAGTAGCGCCAGAGGGCTGGATATTGAAAAATGTG 60
QY	1841	CAACATGTATCAATTTGATCTTCCCTTACCATTGATGAAATATGTCATCGAATTGGG 1900
Db	61	CAACATGTTATCAATTTGATCTTCCCTTACCATTGATGAAATATGTCATCGAATTGGG 120
QY	1901	CGTACTGCTGCTGTGTGGAACTACTGCAGAGCAATTCCTTTTGTGATCTTGAATCGAT 1960
Db	121	CGTACTGCTGCTGTGTGGAACTACTGCAGAGCAATTCCTTTTGTGATCTTGAATCGAT 180
QY	1961	AACCATTTAGCACAGCCTCTAGTAAAAGTATTGACAGATG 2000
Db	181	AACCATTTAGCACAGCCTCTAGTAAAAGTATTGACAGATG 220
<hr/>		
RESULT 11		
ABL97687	ID	ABL97687 standard; DNA; 2194 BP.
XX	AC	ABL97687;
XX	DT	21-JUN-2002 (first entry)
DE		Human testicular antigen encoding DNA fragment SEQ ID NO: 2339.
KM		Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW		reproductive system disorder; urinary system disorder; gene therapy;
KX		cardiovascular disorder; respiratory disorder; neurological disorder;
KX		gastrointestinal disease; infection; cytostatic; gene; ds.
OS		Homo sapiens.
PN		WO200155317-A2.
PD		02-AUG-2001.
PF		17-JAN-2001; 2001WO-US001329.
PR		31-JAN-2000; 2000US-0179065P.
PR		04-FEB-2000; 2000US-0180628P.
PR		24-FEB-2000; 2000US-0184664P.
PR		02-MAR-2000; 2000US-0186350P.
PR		16-MAR-2000; 2000US-0189874P.
PR		17-MAR-2000; 2000US-0190076P.
PR		18-APR-2000; 2000US-0198123P.
PR		19-MAY-2000; 2000US-020515P.
PR		07-JUN-2000; 2000US-0209467P.
PR		28-JUN-2000; 2000US-0214886P.
PR		30-JUN-2000; 2000US-0215135P.

Sequence 2194 BP; 624 A; 349 C; 400 G; 821 T; 0 U; 0 Other;

Query Match 9.9%; Score 220; DB 4; Length 2194;
Best Local Similarity 100.0%; Pred. No. 3,9e-98;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1781 AAGTCCAGTCTTCTTGTCTACTTCACTAGCTGCCAGAGGGCTGGATATTGAAATGTG 1840
DB 1 AAGTCCAGTCTTCTTGTCTACTTCACTAGCTGCCAGAGGGCTGGATATTGAAATGTG 60
QY 1841 CACATGTTATCAATTTGATCTTCTCTACCTGATGATGATATGTTCAATCGAATGGG 1900
DB 61 CACATGTTATCAATTTGATCTTCTCTACCTGATGATGATATGTTCAATCGAATGGG 120
QY 1901 CGTACTGTCGTCGTTGGGAACTGCGACAGCAATTCCTTTTGTGATCTTGAAATCGAT 1960
DB 121 CGTACTGTCGTCGTTGGGAACTGCGACAGCAATTCCTTTTGTGATCTTGAAATCGAT 180
QY 1961 AACCATTTAGCAGAGCCCTCTAGTAAAGTATTGACAGATG 2000
DB 181 AACCATTTAGCAGAGCCCTCTAGTAAAGTATTGACAGATG 220

RESULT 12

AL04794 standard; DNA; 2195 BP.

21-NOV-2001 (first entry)

Human reproductive system related antigen DNA SEQ ID NO: 7482.

Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy; ds.

Homo sapiens.

W0200155320-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001339.

31-JAN-2000; 2000US-0179065P.
04-FEB-2000; 2000US-0180628P.
24-FEB-2000; 2000US-0184664P.
02-MAR-2000; 2000US-0186350P.
16-MAR-2000; 2000US-0189874P.
17-MAR-2000; 2000US-0190076P.
18-APR-2000; 2000US-0198123P.
19-MAY-2000; 2000US-0205515P.
07-JUN-2000; 2000US-0209467P.
28-JUN-2000; 2000US-0214886P.
30-JUN-2000; 2000US-0215135P.
07-JUL-2000; 2000US-0216647P.
11-JUL-2000; 2000US-0216880P.
11-JUL-2000; 2000US-0217487P.
14-JUL-2000; 2000US-0217496P.
26-JUL-2000; 2000US-0218290P.
26-JUL-2000; 2000US-0220963P.
26-JUL-2000; 2000US-0220964P.
14-AUG-2000; 2000US-0224518P.
14-AUG-2000; 2000US-0224519P.
14-AUG-2000; 2000US-0225213P.
14-AUG-2000; 2000US-0225214P.
14-AUG-2000; 2000US-0225266P.
14-AUG-2000; 2000US-0225267P.
14-AUG-2000; 2000US-0225268P.
14-AUG-2000; 2000US-0225270P.
14-AUG-2000; 2000US-0225447P.
14-AUG-2000; 2000US-0225757P.
14-AUG-2000; 2000US-0225758P.
14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-0226279P.
22-AUG-2000; 2000US-0226681P.
22-AUG-2000; 2000US-0226686P.
22-AUG-2000; 2000US-0227182P.
23-AUG-2000; 2000US-0227009P.
30-AUG-2000; 2000US-0228924P.
01-SEP-2000; 2000US-0229287P.
01-SEP-2000; 2000US-0229343P.
01-SEP-2000; 2000US-0229344P.
01-SEP-2000; 2000US-0229345P.
05-SEP-2000; 2000US-0229509P.
05-SEP-2000; 2000US-0229513P.
06-SEP-2000; 2000US-0230437P.
06-SEP-2000; 2000US-0230438P.
08-SEP-2000; 2000US-0231242P.
08-SEP-2000; 2000US-0231243P.
08-SEP-2000; 2000US-0231244P.
08-SEP-2000; 2000US-0231413P.
08-SEP-2000; 2000US-0231414P.
08-SEP-2000; 2000US-0232080P.
08-SEP-2000; 2000US-0232081P.
12-SEP-2000; 2000US-0231968P.
14-SEP-2000; 2000US-0232397P.
14-SEP-2000; 2000US-0232398P.
14-SEP-2000; 2000US-0232399P.
14-SEP-2000; 2000US-0232400P.
14-SEP-2000; 2000US-0232401P.
14-SEP-2000; 2000US-0233063P.
14-SEP-2000; 2000US-0233064P.
14-SEP-2000; 2000US-0233065P.
21-SEP-2000; 2000US-0234223P.
21-SEP-2000; 2000US-0234274P.
25-SEP-2000; 2000US-0234997P.
25-SEP-2000; 2000US-0234998P.
26-SEP-2000; 2000US-0235484P.
27-SEP-2000; 2000US-0235834P.
27-SEP-2000; 2000US-0235836P.
29-SEP-2000; 2000US-0236227P.
29-SEP-2000; 2000US-0236367P.
29-SEP-2000; 2000US-0236368P.
29-SEP-2000; 2000US-0236369P.
29-SEP-2000; 2000US-0236370P.
02-OCT-2000; 2000US-0236802P.
02-OCT-2000; 2000US-0237037P.
02-OCT-2000; 2000US-0237038P.
02-OCT-2000; 2000US-0237039P.
02-OCT-2000; 2000US-0237040P.
13-OCT-2000; 2000US-0239335P.
13-OCT-2000; 2000US-0239337P.
20-OCT-2000; 2000US-0240960P.
20-OCT-2000; 2000US-0241221P.
20-OCT-2000; 2000US-0241786P.
20-OCT-2000; 2000US-0241788P.
20-OCT-2000; 2000US-0241787P.
20-OCT-2000; 2000US-0241808P.
20-OCT-2000; 2000US-0241809P.
20-OCT-2000; 2000US-0241826P.
01-NOV-2000; 2000US-0244617P.
08-NOV-2000; 2000US-0246474P.
08-NOV-2000; 2000US-0246475P.
08-NOV-2000; 2000US-0246476P.
08-NOV-2000; 2000US-0246477P.
08-NOV-2000; 2000US-0246478P.
08-NOV-2000; 2000US-0246523P.
08-NOV-2000; 2000US-0246524P.
08-NOV-2000; 2000US-0246525P.
08-NOV-2000; 2000US-0246526P.
08-NOV-2000; 2000US-0246527P.
08-NOV-2000; 2000US-0246528P.
08-NOV-2000; 2000US-0246532P.
08-NOV-2000; 2000US-0246603P.
08-NOV-2000; 2000US-0246619P.
08-NOV-2000; 2000US-0246611P.

RESULT 13	
ABL97688	
ID	ABL97688 standard; DNA; 2195 BP.
XX	
AC	ABL97688;
XX	
DT	21-JUN-2002 (first entry)
DE	Human testicular antigen encoding DNA fragment SEQ ID NO: 2340.
XX	
KM	Human; testicular antigen; testes; cancer; metastasis; immune disorder
KM	reproductive system disorder; urinary system disorder; gene therapy;
KM	cardiovascular disorder; respiratory disorder; neurological disorder;
KM	gastrointestinal disease; infection; cytostatic; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200155317-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US001329.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214866P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216800P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217496P.
PR	14-JUL-2000; 2000US-0218290P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224516P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225477P.
PR	14-AUG-2000; 2000US-0225757P.
PR	14-AUG-2000; 2000US-0225758P.
PR	14-AUG-2000; 2000US-0225759P.
PR	16-AUG-2000; 2000US-0226273P.
PR	22-AUG-2000; 2000US-0226681P.
PR	22-AUG-2000; 2000US-0226686P.
PR	22-AUG-2000; 2000US-0227182P.
PR	23-AUG-2000; 2000US-0227109P.
PR	30-AUG-2000; 2000US-0228924P.
PR	01-SEP-2000; 2000US-0229287P.
PR	01-SEP-2000; 2000US-0229343P.
PR	01-SEP-2000; 2000US-0229344P.
PR	01-SEP-2000; 2000US-0229345P.
PR	05-SEP-2000; 2000US-0229509P.
PR	05-SEP-2000; 2000US-0229513P.
PR	06-SEP-2000; 2000US-0230437P.
PR	06-SEP-2000; 2000US-0230438P.
PR	08-SEP-2000; 2000US-0231242P.
PR	08-SEP-2000; 2000US-0231243P.
PR	08-SEP-2000; 2000US-0231244P.
PR	08-SEP-2000; 2000US-0231413P.
PR	08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0233080P.
PR 08-SEP-2000; 2000US-0233081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0233397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0233399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-0233401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246509P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251889P.
PR 08-DEC-2000; 2000US-0251900P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM,
DR WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT for preventing, diagnosing and/or treating testicular cancer.
XX
XX Disclosure; SEQ ID NO 2340; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention
XX
SQ Sequence 2195 BP; 624 A; 350 C; 400 G; 821 T; 0 U; 0 Other;
Query Match 9.9%; Score 220; DB 4; Length 2195;
Best Local Similarity 100.0%; Pred. No. 3.9e-98;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1781 AAGTCCAGTCTTGTGCTACTTCACTAGCTGCCAGAGGGCTGATTAAGAAATG 1840
Db 1 AAGTCCAGTCTTGTGCTACTTCACTAGCTGCCAGAGGGCTGATTAAGAAATG 60
QY 1841 CAACATGTTATCAATTTTGATCTTCTTCAACATTAAGAAATGTTCAATGGAATGG 1900
Db 61 CAACATGTTATCAATTTTGATCTTCTTCAACATTAAGAAATGTTCAATGGAATGG 120
QY 1901 CGTACTGTCGTTGGGAATCTGAGAGCAATTTCTTTTGGATCTGAATCGAT 1960
Db 121 CGTACTGTCGTTGGGAATCTGAGAGCAATTTCTTTTGGATCTGAATCGAT 180
QY 1961 AACCATTTAGCACAGCTCTAGTAAAGTATTGACAGATG 2000
Db 181 AACCATTTAGCACAGCTCTAGTAAAGTATTGACAGATG 220
RESULT 14
AL04795
ID AAL04795 standard; DNA; 2197 BP.
XX
AC AAL04795;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7483.
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; db.
XX
OS Homo sapiens.
XX
FN WO200155320-A2.

Wed Jul 28 11:42:29 2004

XX 02-AUG-2001.
 PD 17-JAN-2001; 2001WO-US001339.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214866P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230488P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232337P.
 PR 14-SEP-2000; 2000US-0232338P.
 PR 14-SEP-2000; 2000US-0232339P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233663P.
 PR 14-SEP-2000; 2000US-0233664P.
 PR 14-SEP-2000; 2000US-0233665P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249266P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX

DR WPI: 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure: SEQ ID NO 7483; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 2197 BP; 624 A; 349 C; 401 G; 823 T; 0 U; 0 Other;

Query Match 9.9%; Score 220; DB 4; Length 2197;
Best Local Similarity 100.0%; Pred. No. 3,9e-98;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1781 AAGTCCAGTTCTGTGTGCTACTCTAGTACGCTCCAGAGGCGGATTTGAAATGTG 1840
DB 1 AAGTCCAGTTCTGTGTGCTACTCTAGTACGCTCCAGAGGCGGATTTGAAATGTG 60

CY 1841 CAACATGTTATCAATTTTGATCTTCCTTACCATTTGATGATATATGTTATCGAATTGGG 1900
DB 61 CAACATGTTATCAATTTTGATCTTCCTTACCATTTGATGATATATGTTATCGAATTGGG 120

CY 1901 CGTCTGCTGCTGTGTGGGAATACGCGAGAGCAATTCCTTTTGTATCTTGATCGGAT 1960
DB 121 CGTCTGCTGCTGTGTGGGAATACGCGAGAGCAATTCCTTTTGTATCTTGATCGGAT 180

CY 1961 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000
DB 181 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 15
ABL97689
ID ABL97689 standard; DNA; 2197 BP.
XX
XX ABL97689;
AC
XX
XX 21-JUN-2002 (first entry)
DT
XX
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2341.
DE
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KM reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestnal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX WO20015317-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001329.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225478P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227709P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0235837P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249247P.
 PR 17-NOV-2000; 2000US-0249248P.
 PR 17-NOV-2000; 2000US-0249249P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPL; 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides, useful
 PT for preventing, diagnosing and/or treating testicular cancer.
 XX
 PS Disclosure; SEQ ID NO 2341; 766bp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a DNA encoding a
 CC protein fragment of the invention
 XX
 SQ Sequence 2197 BP; 624 A; 349 C; 401 G; 823 T; 0 U; 0 Other;

Query March 9.9%; Score 220; DB 4; Length 2197;
 Best Local Similarity 100.0%; Pred. No. 3.9e-98;

Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1781 AAGTCCAGTCTTGTGCTACTTCACTGCTGAGTCCAGAGGCTGGATATTGAAAATGTG 1840
 DB 1 AAGTCCAGTCTTGTGCTACTTCACTGCTGAGTCCAGAGGCTGGATATTGAAAATGTG 60
 QY 1841 CAACATGTATCAATTTTGAATCTTCTTCACTGATGATGAATATGTCATCGAATTTGG 1900
 DB 61 CAACATGTATCAATTTTGAATCTTCTTCACTGATGATGAATATGTCATCGAATTTGG 120
 QY 1901 CGTACTGATGCTGTGGGAATACGTGACAGACAAATTCCTTTTGTGATCTTGAATCGGAT 1960
 DB 121 CGTACTGATGCTGTGGGAATACGTGACAGACAAATTCCTTTTGTGATCTTGAATCGGAT 180
 QY 1961 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000
 DB 181 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 220
 Search completed: July 27, 2004, 14:09:31
 Job time : 885 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 13:19:58 ; Search time 166 Seconds

(Without alignments)
7435.005 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224

Sequence: 1 acttgagtcaccatgagggg.....aagctctgtgttgcata 2224

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

- 1: Issued Patents NA.*
- 2: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	20.1	504	US-09-621-976-2923	Sequence 2923, Ap
2	340	15.3	420	US-08-833-381-1475	Sequence 1475, Ap
3	21	0.9	1830121	US-09-537-884-1	Sequence 1, Appl
4	21	0.9	1830121	US-09-643-990A-1	Sequence 1, Appl
5	19	0.9	2319	US-09-058-489-90	Sequence 90, Appl
6	19	0.9	2451	US-09-976-594-786	Sequence 786, Appl
7	19	0.9	2850	US-09-184-964-2	Sequence 2, Appl
8	19	0.9	3408	US-09-058-489-14	Sequence 14, Appl
9	19	0.9	4416	US-09-058-489-17	Sequence 17, Appl
10	19	0.9	5322	US-09-058-489-13	Sequence 13, Appl
11	19	0.9	536165	US-09-214-808-1	Sequence 1, Appl
12	18	0.8	339	US-08-651-155B-116	Sequence 116, App
13	18	0.8	339	US-09-134-036B-116	Sequence 116, App
14	18	0.8	421	US-09-232-575-116	Sequence 116, App
15	18	0.8	421	US-09-389-681-116	Sequence 116, App
16	18	0.8	421	US-09-620-405B-116	Sequence 116, App
17	18	0.8	421	US-09-339-338-116	Sequence 116, App
18	18	0.8	421	US-09-433-826B-116	Sequence 116, App
19	18	0.8	421	US-09-604-287A-116	Sequence 116, App
20	18	0.8	421	US-09-285-480-116	Sequence 116, App
21	18	0.8	421	US-09-834-759-116	Sequence 116, App
22	18	0.8	532	US-09-318-443-3	Sequence 3, Appl
23	18	0.8	598	US-09-833-381-99	Sequence 99, Appl
24	18	0.8	813	US-09-540-236-226	Sequence 226, App
25	18	0.8	1205	US-09-533-029-5	Sequence 5, Appl
26	18	0.8	1536	US-09-318-443-5	Sequence 5, Appl
27	18	0.8	1682	US-09-318-443-7	Sequence 7, Appl

C	28	18	0.8	1796	3	US-09-255-392-1	Sequence 1, Appl
	29	18	0.8	1876	3	US-09-082-092-5	Sequence 5, Appl
	30	18	0.8	1876	4	US-09-885-722A-5	Sequence 5, Appl
	31	18	0.8	1914	4	US-09-291-922-19	Sequence 19, Appl
C	32	18	0.8	2005	2	US-08-484-200-1	Sequence 1, Appl
C	33	18	0.8	2005	2	US-08-477-493-1	Sequence 1, Appl
C	34	18	0.8	2005	3	US-08-465-375-1	Sequence 1, Appl
C	35	18	0.8	2005	3	US-09-788-070-1	Sequence 1, Appl
C	36	18	0.8	2005	4	US-10-142-373-1	Sequence 1, Appl
C	37	18	0.8	2478	4	US-09-328-352-1848	Sequence 1848, App
C	38	18	0.8	3001	4	US-09-539-333D-151	Sequence 151, App
	39	18	0.8	3103	3	US-08-826-246-3	Sequence 3, Appl
	40	18	0.8	3103	3	US-08-944-495-3	Sequence 3, Appl
	41	18	0.8	3103	3	US-09-126-640-2	Sequence 2, Appl
	42	18	0.8	3103	4	US-08-925-588-3	Sequence 3, Appl
	43	18	0.8	3103	4	US-09-288-292A-2	Sequence 2, Appl
	44	18	0.8	3103	4	US-09-372-044-3	Sequence 3, Appl
	45	18	0.8	3103	4	US-08-825-466-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-621-976-2923
; Sequence 2923 Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Gioberto, S.
; APPLICANT: Gioberto, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2923
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: 224..472
US-09-621-976-2923

Query Match	20.1%;	Score 448;	DB 4;	Length 504;
Best Local Similarity	100.0%;	Pred. No. 2.9e-216;		
Matches 448;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	775	TTAGGACTCCATCTTGGACATTTATCAGACGAGGTAAGTTCGACAAATACGACATAT	834	
DB	1	TGAGACTCCATCTTGGACATTTATCAGACGAGGTAAGTTCGACAAATACGACATAT	60	
QY	835	TTTGTGGAAGTGTCTGACATGATGACACCAACGAAATTTGACTTTGAAGAGCTTA	894	
DB	61	TTTGTGGAAGTGTCTGACATGATGACACCAACGAAATTTGACTTTGAAGAGCTTA	120	
QY	895	TCTGTGAGACACGATTAACAAATGCTTAAGCTGTATCTAGCTTACCTCTGT	954	
DB	121	TCTGTGAGACACGATTAACAAATGCTTAAGCTGTATCTAGCTTACCTCTGT	180	
QY	955	GCAAAATACAGTATTCATCATCTTGACAGACGAGATTGATGCTTGCTCAAC	1014	
DB	181	GCAAAATACAGTATTCATCATCTTGACAGACGAGATTGATGCTTGCTCAAC	240	
QY	1015	AGGCTCTGGAAGAGTGGGCTTTTCTCTACCAATTTGGCTCATATGATGATGATG	1074	
DB	241	AGGCTCTGGAAGAGTGGGCTTTTCTCTACCAATTTGGCTCATATGATGATGATG	300	
QY	1075	AATACTCCAGTCTTTTAAAGAGTTCGAGAAACGAGATGATATATGACACCAAC	1134	
DB	301	AATACTCCAGTCTTTTAAAGAGTTCGAGAAACGAGATGATATATGACACCAAC	360	

QY 1135 TCAGAGATTGTCACACAGATTATTGGAAGCCAGAAAATTTCTTTGGGACTTGCT 1194
DB 361 TCAGAGATTGTCACACAGATTATTGGAAGCCAGAAAATTTCTTTGGGACTTGCT 420
QY 1195 AAGAGCTGTTGTTATATATGCGGGAACC 1222
DB 421 AAGAGCTGTTGTTATATATGCGGGAACC 448

RESULT 2
US-09-833-381-1475
Sequence 1475, Application US/09833381
Patent No. 6672186
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
FILE REFERENCE: 5800-119
CURRENT FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 09/516,448
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 2050
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1475
LENGTH: 420
TYPE: DNA
ORGANISM: Homo sapiens
US-09-833-381-1475

Query Match 15.3%; Score 340; DB 4; Length 420;
Best Local Similarity 99.7%; Pred. No. 1,1e-161;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 820 CAAATACGACATATTCTTGTGGAAGTCTGGAATGATGACACCCAGCAATTTCTGC 879
DB 30 CAAATACGACATATTCTTGTGGAAGTCTGGAATGATGACACCCAGCAATTTCTGC 89
QY 880 TTTTGAAGAGCTTAATCTCTGTCAGACACTGATTAACAATCTGTTAAAGCTGTTATAC 939
DB 90 TTTTGAAGAGCTTAATCTCTGTCAGACACTGATTAACAATCTGTTAAAGCTGTTATAC 149
QY 940 TAAGCTTACTCTGTCGCAAAATACAGTATTCCTATCAATCTGTCAGAGAGATTGAT 999
DB 150 TAAGCTTACTCTGTCGCAAAATACAGTATTCCTATCAATCTGTCAGAGAGATTGAT 209
QY 1000 GGGCTTGCTCAACAGGCTCTGGAAGACTGGGGCTTTCTCTCAATTTGGCTCA 1059
DB 210 GGGCTTGCTCAACAGGCTCTGGAAGACTGGGGCTTTCTCTCAATTTGGCTCA 269
QY 1060 TATGATCATGATGATTAATCTCCAGTCTTTTAAAGATTGACAGAAACAGAGTAT 1119
DB 270 TATGATCATGATGATTAATCTCCAGTCTTTTAAAGATTGACAGAAACAGAGTAT 329
QY 1120 TATTGTAGCAACCACTGGAATTTGTCACACAGATTATTGGAAGCCAGAAAATTTTC 1179
DB 330 TATTGTAGCAACCACTGGAATTTGTCACACAGATTATTGGAAGCCAGAAAATTTTC 389
QY 1180 TTTTGGAGCTTGTTAGAGCTGTTGTATA 1210
DB 390 TTTTGGAGCTTGTTAGAGCTGTTGTATA 420

RESULT 3
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: PB186P3
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 0.9%; Score 21; DB 4; Length 1830121;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1892 CGAATGGGCGTACTGCTGT 1912
DB 261968 CGAATGGGCGTACTGCTGT 261988

RESULT 4

US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams

Owen White
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000

```

;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          0.9% Score 21; DB 4; Length 1830121;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1892 CGAATTGGCGCTACTGTCGT 1912
DB      261968 CGAATTGGCGCTACTGTCGT 261988

RESULT 5
US-09-058-489-90
; Sequence 90, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-089A
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 90
; LENGTH: 2319
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-90

Query Match          0.9% Score 19; DB 3; Length 2319;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1843 ACATGTTATCAATTTGAT 1861
DB      1598 ACATGTTATCAATTTGAT 1616

RESULT 6
US-09-976-594-786
; Sequence 786, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
```

```

;
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 786
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6673549 035282CB1
US-09-976-594-786

Query Match          0.9% Score 19; DB 4; Length 2451;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1843 ACATGTTATCAATTTGAT 1861
DB      1572 ACATGTTATCAATTTGAT 1590

RESULT 7
US-09-184-964-2
; Sequence 2, Application US/09184964
; Patent No. 6391574
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Boyatchuk, Victor L.
; APPLICANT: Ashby, Matthew N.
; TITLE OF INVENTION: APCI AND RCEL: ISOPRENATED CMAX
; TITLE OF INVENTION: PROCESSING ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson P.C.
; STREET: 2200 Sand Hill road, suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/184,964
; FILING DATE: 03-NOV-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/902,774
; FILING DATE: 30-JUL-1997
; APPLICATION NUMBER: 60/023,491
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Syat, Reginald J.
; REGISTRATION NUMBER: 28,172
; REFERENCE/DOCKET NUMBER: 09272-006004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/322-5070
; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic DNA
US-09-184-964-2

Query Match          0.9% Score 19; DB 4; Length 2850;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1687 TTTCTTTTGTCAAGAAAA 1705
      |||||||
DB      919 TTTCTTTTGTCAAGAAAA 937

RESULT 8
US-09-058-489-14
; Sequence 14, Application US/09058489
; Patent No. 610386
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 3408
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-14

Query Match      0.9%; Score 19; DB 3; Length 3408;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1842 AACATGTTATCAATTGGA 1860
      |||||||
DB      2388 AACATGTTATCAATTGGA 2406

RESULT 9
US-09-058-489-17
; Sequence 17, Application US/09058489
; Patent No. 610386
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 4416
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-17

Query Match      0.9%; Score 19; DB 3; Length 4416;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1843 ACATGTTATCAATTGAT 1861
      |||||||
DB      1598 ACATGTTATCAATTGAT 1616

RESULT 10
US-09-058-489-13
; Sequence 13, Application US/09058489

QY      1687 TTTCTTTTGTCAAGAAAA 1705
      |||||||
DB      919 TTTCTTTTGTCAAGAAAA 937

Patent No. 610386
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 5322
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-13

Query Match      0.9%; Score 19; DB 3; Length 5322;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1842 AACATGTTATCAATTGGA 1860
      |||||||
DB      2388 AACATGTTATCAATTGGA 2406

RESULT 11
US-09-214-808-1
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; FILE REFERENCE: Plasmid
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-214-808-1

Query Match      0.9%; Score 19; DB 4; Length 536165;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      404 GCGCGCTTCGAGATCGAA 422
      |||||||
DB      146504 GCGCGCTTCGAGATCGAA 146522

RESULT 12
US-08-651-155B-116
; Sequence 116, Application US/08651155B
; Patent No. 6363401
; GENERAL INFORMATION:
; APPLICANT: Mahan Dr., Michael J.
; APPLICANT: Conner Mr., Christopher P.
; APPLICANT: Hiethoff Mr., Douglas M.
; TITLE OF INVENTION: METHOD AND PROBS FOR THE IDENTIFICATION
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
```

TITLE OF INVENTION: INFECTION
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chairman, Byrum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen M., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-651-155B-116

Query Match 0.8%; Score 18; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTCGAGCTTCATCATCAG 140
Db 184 CTCGAGCTTCATCATCAG 201

RESULT 13
US-09-194-036B-116
Sequence 116, Application US/09194036B
Patent No. 6548246
GENERAL INFORMATION:
APPLICANT: Mahan, Michael J.
Conner, Christopher P.
Hiechoff, Douglas M.
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
INFECTION
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Mountain View
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,036B
FILING DATE: 17-NO. 6548246-1998

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/08208
FILING DATE: 1997-05-16
APPLICATION NUMBER: US 08/651,155
FILING DATE: 1996-05-17
ATTORNEY/AGENT INFORMATION:
NAME: Shantanu Basu
REGISTRATION NUMBER: 43,318
REFERENCE/DOCKET NUMBER: 220002060601
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 813-5995
TELEFAX: (650) 494-0792
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: DNA (other)
SEQUENCE DESCRIPTION: SEQ ID NO: 116:
US-09-194-036B-116

Query Match 0.8%; Score 18; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTCGAGCTTCATCATCAG 140
Db 184 CTCGAGCTTCATCATCAG 201

RESULT 14
US-09-222-575-116
Sequence 116, Application US/09222575
Patent No. 6387837
GENERAL INFORMATION:
APPLICANT: Yugu, Jjiang
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
FILE REFERENCE: 210121.470
CURRENT APPLICATION NUMBER: US/09/222,575
CURRENT FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 174
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 116
LENGTH: 421
TYPE: DNA
ORGANISM: Human
US-09-222-575-116

Query Match 0.8%; Score 18; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GGTCGAGAAATGAAGAG 1399
Db 12 GGTCGAGAAATGAAGAG 29

RESULT 15
US-09-389-681-116
Sequence 116, Application US/09389681A
Patent No. 6518237
GENERAL INFORMATION:
APPLICANT: Yugu, Jjiang
APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ. ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 116
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-389-681-116

Query Match 0.8%; Score 18; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GGTCCAGAAATGAAGAAG 1399
|||
Db 12 GGTCCAGAAATGAAGAAG 29

Search completed: July 27, 2004, 18:11:16
Job time : 176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 16:33:16 ; Search time 999 Seconds
(without alignments)
10882.460 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224
Sequence: 1 acttgagtcaccatggggg999.....aagtcgtgtgttgatgca 2224

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3216467 seqs, 2444149694 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PC07_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PC07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1288	57.9	1984	US-10-104-047-1652
2	340	15.3	420	US-09-833-381-1475
3	278	12.3	591	US-09-764-891-1170
4	273	12.3	539	US-10-029-386-12207
5	271	12.2	271	US-10-029-386-25907
6	220	9.9	2194	US-09-764-891-7481
7	220	9.9	2195	US-09-764-891-7482
8	220	9.9	2197	US-09-764-891-7483
9	129	5.8	516	US-10-029-386-8746
10	128	5.8	153	US-10-029-386-22446
11	76	3.4	570	US-10-027-632-201125
12	76	3.4	16	US-10-027-632-201125
13	60	2.7	10	US-09-808-975-15365
14	36	1.6	1383	US-10-282-122A-41827

15	25	1.1	1335	16	US-10-369-493-24127	Sequence 24127, A
16	24	1.1	1347	16	US-10-369-493-24036	Sequence 24036, A
17	23	1.0	13030	16	US-10-408-168-1	Sequence 1, Appli
18	21	0.9	1642	9	US-09-815-243-6893	Sequence 6893, Ap
19	21	0.9	1642	13	US-10-282-122A-21926	Sequence 21926, A
20	21	0.9	1830121	15	US-10-329-960-1	Sequence 1, Appli
21	21	0.9	1830121	16	US-10-329-960-1	Sequence 1, Appli
22	20	0.9	620	13	US-10-027-632-8290	Sequence 8290, Ap
23	20	0.9	620	13	US-10-027-632-8290	Sequence 8290, Ap
24	20	0.9	620	16	US-10-027-632-8290	Sequence 8290, Ap
25	20	0.9	620	16	US-10-027-632-8290	Sequence 8290, Ap
26	20	0.9	633	13	US-10-027-632-22814	Sequence 22814, A
27	20	0.9	633	16	US-10-027-632-22814	Sequence 22814, A
28	20	0.9	637	13	US-10-027-632-214123	Sequence 214123, A
29	20	0.9	637	16	US-10-027-632-214123	Sequence 214123, A
30	20	0.9	704	13	US-10-027-632-98678	Sequence 98678, A
31	20	0.9	704	16	US-10-027-632-98678	Sequence 98678, A
32	20	0.9	710	13	US-10-027-632-163741	Sequence 163741, A
33	20	0.9	710	16	US-10-027-632-163741	Sequence 163741, A
34	20	0.9	1193	13	US-10-424-539-14492	Sequence 14492, A
35	20	0.9	2763	17	US-10-437-963-69857	Sequence 69857, A
36	20	0.9	2934	17	US-10-437-963-69859	Sequence 69859, A
37	20	0.9	9832	15	US-10-311-455-630	Sequence 630, Appl
38	20	0.9	96595	12	US-09-997-722-43	Sequence 43, Appl
39	20	0.9	122859	13	US-10-087-132-37	Sequence 37, Appl
40	19	0.9	133	13	US-10-029-386-23676	Sequence 23676, A
41	19	0.9	138	9	US-09-864-761-28835	Sequence 28835, A
42	19	0.9	367	13	US-10-424-539-16572	Sequence 16572, A
43	19	0.9	418	10	US-09-918-995-35717	Sequence 35717, A
44	19	0.9	510	13	US-10-027-632-69258	Sequence 69258, A
45	19	0.9	510	15	US-10-029-386-9976	Sequence 9976, Ap

ALIGNMENTS

RESULT 1
US-10-104-047-1652
; Sequence 1652, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1652
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1652

Query Match	57.9%	Score 1288;	DB 16;	Length 1984;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 1488;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	733	AGACCAAAAGTGACCTACATACCCCTCTCTCCACCTGAGATGAGAGCTCATCTTTGC	792	
DB	292	AGACCAAAAGTGACCTACATACCCCTCTCTCCACCTGAGATGAGAGCTCATCTTTGC	351	
QY	793	ACATTATGACAGGAGATTAATCTTGACAAATAGACACTATCTTGGAAGTCTGG	852	
DB	352	ACATTATGACAGGAGATTAATCTTGACAAATAGACACTATCTTGGAAGTCTGG	411	
QY	853	ACATGATGACACGACGACATCTGACTTTGAGAGACTATCTCTGACACACTGAA	912	
DB	412	ACATGATGACACGACGACATCTGACTTTGAGAGACTATCTCTGACACACTGAA	471	
QY	913	TAAACAATGTGTAAGCTGTATATACATAGCTTCTGTGCAAAATACAGTATTC	972	

Db 472 TAAACATTCGTAAGAGCTGTTATCTAGCTTACCTCGTCCAAAAATACAGTATTC 531
Qy 973 TATCATCTTGCAGAGAGATTTGATGGCTTGCTCAAAACAGGGCTCGGAGAGCTGC 1032
Db 532 TATCATCTTGCAGAGAGATTTGATGGCTTGCTCAAAACAGGGCTCGGAGAGCTGC 591
Qy 1033 GGGCTTTCTCCTCAACAAATTTGGCTCATATGATGATGAAATTAATGCGAGTGT 1092
Db 592 GGGCTTTCTCCTCAACAAATTTGGCTCATATGATGATGAAATTAATGCGAGTGT 651
Qy 1093 TAAAGAGTGCAGAGAACAGAGTATTTATTTAGACCAACAGCGAATTTGGTCAACA 1152
Db 652 TAAAGAGTGCAGAGAACAGAGTATTTATTTAGACCAACAGCGAATTTGGTCAACA 711
Qy 1153 GATTATTTGAAAGCCAGAAATTTCTTTGGAGCTGTGTAAAGAGCTGTGTATATA 1212
Db 712 GATTATTTGAAAGCCAGAAATTTCTTTGGAGCTGTGTAAAGAGCTGTGTATATA 771
Qy 1213 TGGGGGAAACCCAGCTGGAGATTCATTTGACAAATATGTAAGAGCTGTATATATTA 1272
Db 772 TGGGGGAAACCCAGCTGGAGATTCATTTGACAAATATGTAAGAGCTGTATATATTA 831
Qy 1273 TGCTACTCTGGAAGACTGATGATATCATAGGCAAAAGAAAGATTGCTCAACAGAT 1332
Db 832 TGCTACTCTGGAAGACTGATGATATCATAGGCAAAAGAAAGATTGCTCAACAGAT 891
Qy 1333 CAATATCTTACTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392
Db 892 CAATATCTTACTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
Qy 1393 GAAGAGTTATTTCTGCGCCAGAGATGCTCAAGAAAGAAAGAAAGAAAGAAAGAAAG 1452
Db 952 GAAGAGTTATTTCTGCGCCAGAGATGCTCAAGAAAGAAAGAAAGAAAGAAAGAAAG 1011
Qy 1453 CAGTGCAACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAGATTTTAAAGTCAATTA 1512
Db 1012 CAGTGCAACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAGATTTTAAAGTCAATTA 1071
Qy 1513 TCGCTTTGCTGCTGTTGGAAGAAGGGTGGAGATGATGATGATGATGATGATGATGAT 1572
Db 1072 TCGCTTTGCTGCTGTTGGAAGAAGGGTGGAGATGATGATGATGATGATGATGATGAT 1131
Qy 1573 CCAAGTTGGCCAGTTCTCAAAAGAAAGAAAGCTGTTGAATTTCTCGAAACATAGGGGA 1632
Db 1132 CCAAGTTGGCCAGTTCTCAAAAGAAAGAAAGCTGTTGAATTTCTCGAAACATAGGGGA 1191
Qy 1633 TGAAGAAGCTATGCTTTTGTGGAAGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1692
Db 1192 TGAAGAAGCTATGCTTTTGTGGAAGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1251
Qy 1693 TTGTCAAGAAAAATATCACTACAGATATCATGATGATGATGATGATGATGATGATGAT 1752
Db 1252 TTGTCAAGAAAAATATCACTACAGATATCATGATGATGATGATGATGATGATGATGAT 1311
Qy 1753 GCAAGCTCTTGGAGATTTTGGCTTTGGAAGTGGCCAGCTTCTTGTCTTCACTCAAG 1812
Db 1312 GCAAGCTCTTGGAGATTTTGGCTTTGGAAGTGGCCAGCTTCTTGTCTTCACTCAAG 1371
Qy 1813 TGCCAGAGGGCTGATATTTGAAAAATGCAACATGATATCAATTTTGAATTTCTTCTAC 1872
Db 1372 TGCCAGAGGGCTGATATTTGAAAAATGCAACATGATATCAATTTTGAATTTCTTCTAC 1431
Qy 1873 CATTGATGAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1932
Db 1432 CATTGATGAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491
Qy 1933 AATTTCCTTTTGAATCTGTAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1992
Db 1492 AATTTCCTTTTGAATCTGTAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1551
Qy 1993 GACAGATGCTCAACAGAGATGTTCTGATGATGATGATGATGATGATGATGATGAT 2052
Db 2052 GACAGATGCTCAACAGAGATGTTCTGATGATGATGATGATGATGATGATGATGAT

Db 1552 GACAGATGCTCAACAGAGATGTTCTGATGATGATGATGATGATGATGATGATGAT 1611
Qy 2053 TCCTGGCTTCAGTGGTATGATCAAGAGAAACGTTTGGATTCAGTTGATATCCAGAAAGG 2112
Db 1612 TCCTGGCTTCAGTGGTATGATCAAGAGAAACGTTTGGATTCAGTTGATATCCAGAAAGG 1671
Qy 2113 CAAGAGACTTTGAACACAGCTGGCTTTTCTTCTTCAAGAGCTCCCAATCAGATGATGA 2172
Db 1672 CAAGAGACTTTGAACACAGCTGGCTTTTCTTCTTCAAGAGCTCCCAATCAGATGATGA 1731
Qy 2173 TGAGTCATGGGATTTAAAGCCAAACATCCTTCAAGTGTGTGTTGATGCA 2224
Db 1732 TGAGTCATGGGATTTAAAGCCAAACATCCTTCAAGTGTGTGTTGATGCA 1783

RESULT 2

US-09-833-381-1475
; Sequence 1475, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1475
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1475

Query Match 15.3%; Score 340; DB 9; Length 420;
Best Local Similarity 99.7%; Pred. No. 4.9e-11;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 820 CAATATACACACTATTTCTTGTGGAAGTGTGACATGATGACACACAGCAATTCAGAC 879
Db 30 CAATATACACACTATTTCTTGTGGAAGTGTGACATGATGACACACAGCAATTCAGAC 89
Qy 880 TTTGGAAGACTATCTCTGTGACAGACTGATATTAACAACATTTGCTTAAAGCTGTATAC 939
Db 90 TTTGGAAGACTATCTCTGTGACAGACTGATATTAACAACATTTGCTTAAAGCTGTATAC 149
Qy 940 TAAGCTTACTCTGTGCAAAAATACAGATTCATATCACTTGCAGAGAGATTTGAT 999
Db 150 TAAGCTTACTCTGTGCAAAAATACAGATTCATATCACTTGCAGAGAGATTTGAT 209
Qy 1000 GGGTTGTGCTCAACAGGGTCTGGGAAGACTGGGGCTTTTCTCTACCAATTTGGCTCA 1059
Db 210 GGGTTGTGCTCAACAGGGTCTGGGAAGACTGGGGCTTTTCTCTACCAATTTGGCTCA 269
Qy 1060 TATGATGATGATGGAATTAATGCGCAGTCTTTAAAGAGTTGAGGAAACAGAGTGTAT 1119
Db 270 TATGATGATGATGGAATTAATGCGCAGTCTTTAAAGAGTTGAGGAAACAGAGTGTAT 329
Qy 1120 TATGTAGCAACCACTCGAGATTTGTCACACAGATTTATTTGGAAGCCAGAAATTTTC 1179
Db 330 TATGTAGCAACCACTCGAGATTTGTCACACAGATTTATTTGGAAGCCAGAAATTTTC 389
Qy 1180 TTTTGGGACTGTGTGAAGAGCTGTGTTATTA 1210
Db 390 TTTTGGGACTGTGTGAAGAGCTGTGTTATTA 420

RESULT 3

US-09-764-891-1170
; Sequence 1170, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1170
LENGTH: 591
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-1170

Query Match 12.5%; Score 278; DB 10; Length 591;
Best Local Similarity 99.7%; Pred. No. 9.3e-138;
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1896 TTGGGCGTACTGTCGTTGGGAATCTGCGACAGCAATTTCTTTTGTGATCTTGAT 1955
Db 30 TTGGGCGTACTGTCGTTGGGAATCTGCGACAGCAATTTCTTTTGTGATCTTGAT 89
QY 1956 CGGATTAACATTTAGCACAGCCTCTAGTAAAGTATTGACAGATGCTCAACAGATGTC 2015
Db 90 CGGATTAACATTTAGCACAGCCTCTAGTAAAGTATTGACAGATGCTCAACAGATGTC 149
QY 2016 CTGCATGTTGGAGAAATTCCTTTAGTACATTCCTGCGCTTCACTGATGATGATCA 2075
Db 150 CTGCATGTTGGAGAAATTCCTTTAGTACATTCCTGCGCTTCACTGATGATGATCA 209
QY 2076 GAGGAAAAGTGTTCATCACTGATGATGATGATGATGATGATGATGATGATGATG 2135
Db 210 GAGGAAAAGTGTTCATCACTGATGATGATGATGATGATGATGATGATGATGATG 269
QY 2136 GGTTCCTTCCTTCGAGGCTCCCAATCCAGTAGTATGATGATGATGATGATGATGAT 2195
Db 270 GGTTCCTTCCTTCGAGGCTCCCAATCCAGTAGTATGATGATGATGATGATGATGAT 329
QY 2196 ACATCCTTCAAGTGTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 2224
Db 330 ACATCCTTCAAGTGTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 358

RESULT 4
US-10-029-386-12207/c
Sequence 12207, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEWICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12207
LENGTH: 539
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC016635.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52
OTHER INFORMATION: EST HUMAN HIT: AA399611.1, EVALUATE 1.00e-64
OTHER INFORMATION: NT HIT: g116156979, EVALUATE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q61496, EVALUATE 5.00e-45
US-10-029-386-12207

Query Match 12.3%; Score 273; DB 15; Length 539;
Best Local Similarity 100.0%; Pred. No. 4.4e-135;

Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1728 GTGATCGGAAACAGAGAGCGGAGCAAGCTCTTGGAGATTTGCTTTGAAAGTGCC 1787
Db 507 GTGATCGGAAACAGAGAGCGGAGCAAGCTCTTGGAGATTTGCTTTGAAAGTGCC 448
QY 1788 CAGTTCTTGTGCTACTTACTAGTACGTCGAGAGGCTGGAATTTGAAATGCAACATG 1847
Db 447 CAGTTCTTGTGCTACTTACTAGTACGTCGAGAGGCTGGAATTTGAAATGCAACATG 388
QY 1848 TTATCAATTTGATCTTCTTACCATGATGATGATGATGATGATGATGATGATGAT 1907
Db 387 TTATCAATTTGATCTTCTTACCATGATGATGATGATGATGATGATGATGATGATGAT 328
QY 1908 GTGCTGTGGAAATCTGCGACAGCAATTTCTTTTGTGATCTTGATGATGATGATGAT 1967
Db 327 GTGCTGTGGAAATCTGCGACAGCAATTTCTTTTGTGATCTTGATGATGATGATGAT 268
QY 1968 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000
Db 267 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 235

RESULT 5
US-10-029-386-25907/c
Sequence 25907, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEWICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 25907
LENGTH: 271
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC016635.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52
OTHER INFORMATION: EST HUMAN HIT: AA399611.1, EVALUATE 2.00e-64
OTHER INFORMATION: NT HIT: g116156979, EVALUATE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q61496, EVALUATE 2.00e-45
US-10-029-386-25907

Query Match 12.2%; Score 271; DB 15; Length 271;
Best Local Similarity 100.0%; Pred. No. 4.9e-134;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1729 TGATCGGAAACAGAGAGCGGAGCAAGCTCTTGGAGATTTGCTTTGAAAGTGCC 1788
Db 271 TGATCGGAAACAGAGAGCGGAGCAAGCTCTTGGAGATTTGCTTTGAAAGTGCC 212
QY 1789 AGTTCTTGTGCTACTTACTAGTACGTCGAGAGGCTGGAATTTGAAATGCAACATG 1848
Db 211 AGTTCTTGTGCTACTTACTAGTACGTCGAGAGGCTGGAATTTGAAATGCAACATG 152
QY 1849 TATCAATTTGATCTTCTTACCATGATGATGATGATGATGATGATGATGATGATGAT 1908
Db 151 TATCAATTTGATCTTCTTACCATGATGATGATGATGATGATGATGATGATGATGATG 92
QY 1909 TCGTTGTGGAAATCTGCGACAGCAATTTCTTTTGTGATCTTGATGATGATGATGAT 1968
Db 91 TCGTTGTGGAAATCTGCGACAGCAATTTCTTTTGTGATCTTGATGATGATGATGATGAT 32
QY 1969 AGCACAGCCTCTAGTAAAGTATTGACAGAT 1999
|||||

Db 31 AGCAGAGCCTCTAGTAAAGTATTGACAGAT 1

RESULT 6

US-09-764-891-7481
 ; Sequence 7481, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7481
 ; LENGTH: 2194
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-764-891-7481

Query Match 9.9%; Score 220; DB 10; Length 2194;
 Best Local Similarity 100.0%; Pred. No. 1.4e-106; Indels 0; Gaps 0;
 Matches 220; Conservative 0; Mismatches 0;

Qy 1781 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 1840
 Db 1 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 60
 Qy 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 1900
 Db 61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 120
 Qy 1901 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 1960
 Db 121 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 180
 Qy 1961 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 2000
 Db 181 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 7

US-09-764-891-7482
 ; Sequence 7482, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7482
 ; LENGTH: 2195
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-764-891-7482

Query Match 9.9%; Score 220; DB 10; Length 2195;
 Best Local Similarity 100.0%; Pred. No. 1.4e-106; Indels 0; Gaps 0;
 Matches 220; Conservative 0; Mismatches 0;

Qy 1781 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 1840
 Db 1 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 60
 Qy 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 1900
 Db 61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 120

Qy 1901 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 1960
 Db 121 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 180
 Qy 1961 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 2000
 Db 181 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 8

US-09-764-891-7483
 ; Sequence 7483, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7483
 ; LENGTH: 2197
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-764-891-7483

Query Match 9.9%; Score 220; DB 10; Length 2197;
 Best Local Similarity 100.0%; Pred. No. 1.4e-106; Indels 0; Gaps 0;
 Matches 220; Conservative 0; Mismatches 0;

Qy 1781 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 1840
 Db 1 AAGTCCAGTCTTGTGCTACTTCACTAGTCCAGAGGCTGATATTGAAATGTG 60
 Qy 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 1900
 Db 61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTCATCGAATTGGG 120
 Qy 1901 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 1960
 Db 121 CGTACTGCTGTGTGGGAATACTGGCAGAGCAATTTCTTTTGAATCTTGATCGAT 180
 Qy 1961 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 2000
 Db 181 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 9

US-10-028-386-8746/C
 ; Sequence 8746, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: AEMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; PRIORITY FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 8746
 ; LENGTH: 516
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

OTHER INFORMATION: MAP TO AC016639.5
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4


```

QY      1358 GATCGCATGTTGGATATGGGTTTG 1382
          |||||
Db      478 GATCGCATGTTGGATATGGGTTTG 502

```

Wed Jul 28 11:42:29 2004

Search completed: July 27, 2004, 20:40:25
Job time : 1006 secs

us-09-714-865b-1.rnpb

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 12:56:43 ; Search time 5696 Seconds
(without alignments)

11659.672 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224
Sequence: 1 actgaagtcaccatcgggggg.....aagtcgtggtttgatgca 2224

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513269 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estbm:*
3: em_estbn:*
4: em_estbu:*
5: em_estbv:*
6: em_estbp:*
7: em_estbr:*
8: em_estbs:*
9: gb_estb1:*
10: gb_estb2:*
11: gb_estb3:*
12: gb_estb4:*
13: gb_estb5:*
14: em_estfun:*
15: em_estfun1:*
16: em_estfun2:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1829	82.2	2523	11	BC030638 Homo sapi
2	665	29.9	850	14	CD251828 AGENCOURT
3	626	28.1	896	14	CD557531 AGENCOURT
4	616	27.7	682	12	BC718812 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
5	616	27.7	693	12	B1560849
6	612	27.5	798	12	BG717110
7	568	25.5	885	14	CD557160
8	544	24.5	981	14	CD359812
9	539	24.2	868	14	CD358073
10	535	24.1	792	13	BQ429126
11	521	23.4	914	14	CD557137
12	509	22.9	747	12	BG719449
13	480	21.6	657	12	BG717812
14	471	21.2	933	14	CD557865
15	469	21.1	651	12	BG717354
16	441	19.8	822	12	B1464267
17	440	19.8	846	12	BG718634
18	430	19.3	861	12	B1461393
19	422	19.0	830	12	B1462207
20	413	18.6	699	12	BG717822
21	412	18.5	743	12	B1463502
22	382	17.6	841	12	B1458115
23	388	17.4	939	14	CD300339
24	381	17.1	884	12	B1560499
25	361	16.2	918	12	B1561001
26	337	15.2	810	12	B1463846
27	326	14.7	758	12	B1559479
28	321	14.4	482	9	AL042306
29	300	13.5	497	13	BK282689
30	300	13.5	555	9	AA399611
31	269	12.1	555	9	AI217144
32	267	12.0	847	12	B1562624
33	246	11.1	917	12	B1463447
34	245	11.0	805	14	CB230120
35	232	10.9	491	9	AA398976
36	223	10.5	335	9	AA383535
37	225	10.1	347	9	AI953070
38	204	9.2	700	12	BG718861
39	161	7.2	604	12	B1460628
40	155	7.0	300	9	AI337133
41	155	7.0	300	9	AI554417
42	155	7.0	300	10	BBS50307
43	155	7.0	300	10	BBS50448
44	152	6.8	690	14	CD767258
45	137	6.2	360	10	AW449170

ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description
1	1829	82.2	2523	11	BC030638 Homo sapiens
2	665	29.9	850	14	CD251828 AGENCOURT
3	626	28.1	896	14	CD557531 AGENCOURT
4	616	27.7	682	12	BC718812 AGENCOURT

2523 bp mRNA linear HTC 17-DEC-2003
IMAGE:4822525, containing frame-shift errors.

BC030638
LOCUS
DEFINITION
Homo sapiens
ACCESSION
BC030638
VERSION
BC030638.1
KEYWORDS
HTC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2523)
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,J., Shenker,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Parker,A.S., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uebli,T.B., Toshyak,S., Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Wallaby,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,J., Helton,E., Kettelman,M., Maman,A., Rodriguez,S.,

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
JOURNAL
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Gilmwood, J., Schmitt, J., Myers, R.M.,
Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalhus, D.E.,
Scherer, A., Schen, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2523)
Strausberg, R.
Submitted (20-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC). Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcabs-remail.nih.gov
Tissue Procurement: Miklos Palkovics, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadamsystemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAX Plate: 33 Row: d Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9507226
This clone has the following problem: frame shifted.

FEATURES
SOURCE

1. 2523
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4822528"
/tissue_type="Testis"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Query Match 82.2%; Score 1829; DB 11; Length 2523;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2199; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 20 GATCAGATTGGAGAGCAAAATCAACCTCATATGCTTCCATGTTCCATATTGAG 79
DB 104 GATGAGATTGGAGAGCAAAATCAACCTCATATGCTTCCATGTTCCATATTGAG 163
QY 80 AAGGATGCTATTCTGGAGAAATGAGAGCAATTTTAAACGAGCTCCAGCTCATATCA 139
DB 164 AAGGATGCTATTCTGGAGAAATGAGAGCAATTTTAAACGAGCTCCAGCTCATATCA 223
QY 140 GAATGATGATGAGACCTTCTCGAAGAGATCATTTGATGAAAGATTTGCTCTGGG 199
DB 224 GAATGATGATGAGACCTTCTCGAAGAGATCATTTGATGAAAGATTTGCTCTGGG 283
QY 200 CGGAATTTTGAACAAGATGCTGTGATGTAATAGGAGATATATCATTCACAAG 259
DB 284 CGGAATTTTGAACAAGATGCTGTGATGTAATAGGAGATATATCATTCACAAG 343
QY 260 GGTGCTTTTGAAGTGAAGAGTTTGGAAACAGAGGTTTTCACACAGAGGTTTGA 319
DB 344 GGTGCTTTTGAAGTGAAGAGTTTGGAAACAGAGGTTTTCACACAGAGGTTTGA 403
QY 320 GATGCTATAGCTCTGCTTTCTGAGAGAGCTAGTATATGATCTGCAAGATTAATCA 379
DB 404 GATGCTATAGCTCTGCTTTCTGAGAGAGCTAGTATATGATCTGCAAGATTAATCA 463

QY 380 CGAAACAGAGGCTTTTCCAAAGAGAGGCGCTATCGAATGGAATTAATTCAGAGCTTCA 439
DB 464 CGAAACAGAGGCTTTTCCAAAGAGAGGCGCTATCGAATGGAATTAATTCAGAGCTTCA 523
QY 440 GGGCCATACAGAGAGGAGAGAGAGTATTTCCGAGGTTCCGAGAGATTTGCTCTA 499
DB 524 GGGCCATACAGAGAGGAGAGAGTATTTCCGAGGTTCCGAGAGATTTGCTCTA 583
QY 500 GGAAGTCCAAATTAATGACTTAGACCCAGACGATGTATGACGGCACTGTGCTTTT 559
DB 584 GGAAGTCCAAATTAATGACTTAGACCCAGACGATGTATGACGGCACTGTGCTTTT 643
QY 560 GGTCTAGAAACAGATTAATGAGGACAGGTATGATGATCTTCCAAAGCAGAAGT 619
DB 644 GGTCTAGAAACAGATTAATGAGGACAGGTATGATGATCTTCCAAAGCAGAAGT 703
QY 620 GCGAGTGAAGTGAAGAGAGTGTGTACAAAGTTTAATGAAGATTAATGAGGCTCT 679
DB 704 GCGAGTGAAGTGAAGAGAGTGTGTACAAAGTTTAATGAAGATTAATGAGGCTCT 763
QY 680 GGAAGATTTTGGAGTCAAGAGCAGAGAGAGAGAGAGAGTGTGTACTCAAGAGCA 739
DB 764 GGAAGATTTTGGAGTCAAGAGCAGAGAGAGAGAGAGAGTGTGTACTCAAGAGCA 823
QY 740 AAAGTGACCTTACATACCCCTCTCCACCTGAGATGAGACTTCATCTTGGACATTA 799
DB 824 AAAGTGACCTTACATACCCCTCTCCACCTGAGATGAGACTTCATCTTGGACATTA 883
QY 800 CAGACAGGCAATTAATCTTGACAAATAGACACTATCTTGTGAAAGTGTGACATGAT 859
DB 884 CAGACAGGCAATTAATCTTGACAAATAGACACTATCTTGTGAAAGTGTGACATGAT 943
QY 860 GCACACAGCAATTTGACTTGTGAAGAGCTATCTGTGACAGACTGAATTAACAC 919
DB 944 GCACACAGCAATTTGACTTGTGAAGAGCTATCTGTGACAGACTGAATTAACAC 1003
QY 920 ATTGCTAAAGCTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 979
DB 1004 ATTGCTAAAGCTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1063
QY 980 CTTCAGAGAGAGAGTTGATGAGCTGTGCTTCAACAGAGTGTGGAAGAGCTGGCGCTTT 1039
DB 1064 CTTCAGAGAGAGAGTTGATGAGCTGTGCTTCAACAGAGTGTGGAAGAGCTGGCGCTTT 1123
QY 1040 CTTCAGAGAGAGTTGATGAGCTGTGCTTCAACAGAGTGTGGAAGAGCTGGCGCTTT 1099
DB 1124 CTTCAGAGAGAGTTGATGAGCTGTGCTTCAACAGAGTGTGGAAGAGCTGGCGCTTT 1183
QY 1100 TTGCAGAGAGAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1159
DB 1184 TTGCAGAGAGAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1243
QY 1160 TTGCAGAGAGAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1219
DB 1244 TTGCAGAGAGAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1303
QY 1220 ACCGAGCTGGACATTCATTTGACAAATAGTATGACAGCTGTATTAATTAATTAAT 1279
DB 1304 ACCGAGCTGGACATTCATTTGACAAATAGTATGACAGCTGTATTAATTAATTAATTA 1363
QY 1280 COTGGAAGATGATGATATCATAGAGCAAGAAAGATTTGCTCAACAGATCAATAC 1339
DB 1364 CC-GGAAGATGATGATATCATAGAGCAAGAAAGATTTGCTCAACAGATCAATAC 1422
QY 1340 TTAGTTTGAATGAAGCTGATGCTATTTGGATATGAGGTTTGGTCCAGAAATGAAG 1399
DB 1423 TTAGTTTGAATGAAGCTGATGCTATTTGGATATGAGGTTTGGTCCAGAAATGAAG 1482
QY 1400 TTAATTTTCCAGAGAGTCCATTAAGAGAAAGAGGCGCAACCTTATGTTAGTGA 1459
DB 1483 TTAATTTTCCAGAGAGTCCATTAAGAGAAAGAGGCGCAACCTTATGTTAGTGA 1542

QY 1460 ACTTTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1519
Db 1543 ACTTTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1602
QY 1520 GTTGTCTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1579
Db 1603 GTTGTCTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1662
QY 1580 GGCCAGTCTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1639
Db 1663 GGCCAGTCTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1722
QY 1640 ACTATGCTCTTGTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1699
Db 1723 ACTATGCTCTTGTTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1782
QY 1700 GAAAAAATATCACTACAGAGTATCCATGATCGGGAACAGAGAGCGGAGCAAGCT 1759
Db 1783 GAAAAAATATCACTACAGAGTATCCATGATCGGGAACAGAGAGCGGAGCAAGCT 1842
QY 1760 CTTCAGAGTCTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1819
Db 1843 CTTCAGAGTCTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1902
QY 1820 GGCGTGGATATTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1879
Db 1903 GGCGTGGATATTCAGAGGAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1962
QY 1880 GAATATGCTATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1939
Db 1963 GAATATGCTATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2022
QY 1940 TTTTTCATCTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 1999
Db 2023 TTTTTCATCTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2082
QY 2000 GCTCAACAGAGTCTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2059
Db 2083 GCTCAACAGAGTCTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2142
QY 2060 TTCAAGTATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2119
Db 2143 TTCAAGTATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2202
QY 2120 ACTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2179
Db 2203 ACTTTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2262
QY 2180 TGGGATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2224
Db 2263 TGGGATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTCTGTGT 2307

RESULT 2
CD251828 850 bp mRNA 1linear EST 22-MAY-2003
LOCUS AGENCOURT 14206017 NIH MGC 180 Homo sapiens cDNA clone
DEFINITION IMAGE:30383198 5', mRNA sequence.
CD251828
ACCESSION CD251828.1 GI:31012294
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Daniela S. Gerhard, Ph.D.
COMMENT Office of Cancer Genomics
National Cancer Institute / NIH
Ridg 31 Rm10A07 Bethesda, MD 20892

FEATURES
source
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMU at:
<http://image.llnl.gov>
Plate: NDA446 row: k column: 15
High quality sequence stop: 626.
Location/Qualifiers
1..850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30383198"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/note="Torgan: Ttestis; Vector: pCW-SPORE.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."
ORIGIN
Query Match 29.9%; Score 665; DB 14; Length 850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 GATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTCTCCATATTTAG 79
Db 17 GATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTCTCCATATTTAG 76
QY 80 AAGATAGATATTCGAGAAATGAGACAAATTTTAAAGAGCTCCAGCTTCATCATCA 139
Db 77 AAGATAGATATTCGAGAAATGAGACAAATTTTAAAGAGCTCCAGCTTCATCATCA 136
QY 140 GAATGATGATGAGACCTTCGAGAGATATTCATGAAGAGTGGCTCTGAG 199
Db 137 GAATGATGATGAGACCTTCGAGAGATATTCATGAAGAGTGGCTCTGAG 196
QY 200 CGGAATTTTGGAAACAGAGATGCTGATGATTAAGAGAGATATTCACAAATG 259
Db 197 CGGAATTTTGGAAACAGAGATGCTGATGATTAAGAGAGATATTCACAAATG 256
QY 260 GGTGTTTTCAGAGTGGAAAGATTTTGGAAACAGAGTTCCTCAACAGAGTTTGA 319
Db 257 GGTGTTTTCAGAGTGGAAAGATTTTGGAAACAGAGTTCCTCAACAGAGTTTGA 316
QY 320 GATGATGATGATGCTGATTTTCGAGAGATGATTAAGACTGGAAGATATTCACA 379
Db 317 GATGATGATGATGCTGATTTTCGAGAGATGATTAAGACTGGAAGATATTCACA 376
QY 380 CGGAACAGAGGTTTTCAGAGAGCGGCTATGAGATGAGAAATATTCAGAGCTTCA 439
Db 377 CGGAACAGAGGTTTTCAGAGAGCGGCTATGAGATGAGAAATATTCAGAGCTTCA 436
QY 440 GGCGCATTCAGAGAGGTTGAGAGAGTTCGAGAGTGGCGGAGAGTTTGGTGA 499
Db 437 GGCGCATTCAGAGAGGTTGAGAGAGTTCGAGAGTGGCGGAGAGTTTGGTGA 496
QY 500 GGAAGTCCAAATATGACTTACAGAGAGCGGCTATGAGATGAGAGAGCTGTGGCCTTTT 559
Db 497 GGAAGTCCAAATATGACTTACAGAGAGCGGCTATGAGATGAGAGAGCTGTGGCCTTTT 556
QY 560 GGTTCCTAAGAGCCAGTATTAAGTGCACAGTATGAGTACTCTCAAGAGAGAGT 619
Db 557 GGTTCCTAAGAGCCAGTATTAAGTGCACAGTATGAGTACTCTCAAGAGAGAGT 616
QY 620 GGCAGTGAAGTGAAGAGAGTGTTCAGAGGTTTAAAGTGAAGATATTAAGAGCTCT 679
Db 617 GGCAGTGAAGTGAAGAGAGTGTTCAGAGGTTTAAAGTGAAGATATTAAGAGCTCT 676

QY 680 GAAA 684
Db 677 GAAA 681

RESULT 3
CD557531
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

CD557531 896 bp mRNA linear EST 11-JUN-2003
AECNCOURT 14413281 NIH MGC 180 Homo sapiens cDNA clone
IMAGE:30350465 5', mRNA sequence.
CD557531
CD557531.1 GI:31583599
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 896)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM465 row: 3 column: 10
High quality sequence stop: 592.
Location/Qualifiers
1. 896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30350465"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/note="Organ: Testis; Vector: pCMV-Sport6.1; Site: 1; NCI;
Site 2: EcorV (destroyed); Library is oligo-dT primed and
directionally cloned (EcorV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 28.1%; Score 626; DB 14; Length 896;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 919 CATGGCTAAAGCTGGTAACTAGAGCTTACCTCTGCAAAATAACAGTATTCCTATCAT 978
Db 1 CATGGCTAAAGCTGGTAACTAGAGCTTACCTCTGCAAAATAACAGTATTCCTATCAT 60

QY 979 ACTTGAGAGACGAGATTGATGCTGCTGCTCAACAGGCTCGGAGAGACGCGGCTT 1038
Db 61 ACTTGAGAGACGAGATTGATGCTGCTGCTCAACAGGCTCGGAGAGACGCGGCTT 120

QY 1039 TCTCTACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
Db 121 TCTCTACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

QY 1099 GTTGCAGGAAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
Db 181 GTTGCAGGAAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 1159 TTGGAAGCAGAAATTTCTTTGGAGCTGTGTAAGAGCTGTTATATATATGAGGG 1218

Db 241 TTGGAAGCAGAAATTTCTTTGGAGCTGTGTAAGAGCTGTTATATATATGAGGG 300

QY 1219 AACCCAGCTGGACATTCATTAATGACAAATAGTACAGAGCTGTAATATATATGCTAC 1278
Db 301 AACCCAGCTGGACATTCATTAATGACAAATAGTACAGAGCTGTAATATATATGCTAC 360

QY 1279 TCTTGAAGACTGATGATATCATAGGCAAGAAAGATTTGCTCAACAGATCAATA 1338
Db 361 TCTTGAAGACTGATGATATCATAGGCAAGAAAGATTTGCTCAACAGATCAATA 420

QY 1339 CTGAGTTTGGATGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
Db 421 CTGAGTTTGGATGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 1399 GTTAATTTCTTCCCGAAGATGCGATCAAGAGCAAGCGCCAAACCTTATGTTCAAGTGC 1458
Db 481 GTTAATTTCTTCCCGAAGATGCGATCAAGAGCAAGCGCCAAACCTTATGTTCAAGTGC 540

QY 1459 AACTTTTCCAGAGAAATTTCAAAGTTGGCTGCAAGATTTTAAAGTCAATATCTGT 1518
Db 541 AACTTTTCCAGAGAAATTTCAAAGTTGGCTGCAAGATTTTAAAGTCAATATCTGT 600

QY 1519 TGTGCTGTTGACCAATGAGGATGAGCATGTAGATGTTTCAAGAGACCGCTTCCAGAT 1578
Db 601 TGTGCTGTTGACCAATGAGGATGAGCATGTAGATGTTTCAAGAGACCGCTTCCAGAT 660

QY 1579 TGGCAGTTTCAAAA 1595
Db 661 TGGCAGTTTCAAAA 677

RESULT 4
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

BG718234 682 bp mRNA linear EST 08-MAY-2001
602696209F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828298 5',
mRNA sequence.
BG718234
BG718234.1 GI:13997421
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 682)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LRAM10746 row: a column: 03
High quality sequence stop: 682.
Location/Qualifiers
1. 682
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4828298"
/lab_host="DH10B"
/clone_lib="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI
(gtagag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a

primary library enriched for full-length clones and constructed using the Cap-trapper method (Garnicki, in preparation). Library constructed by M. Brownstein (NIH/NHRI, National Institutes of Health). Note: this is a NIH-MGC Library."

ORIGIN

Query Match 27.7%; Score 616; DB 12; Length 682;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATATTGAG 79
67 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATATTGAG 126
80 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCCAGCTCATCATCA 139
127 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCCAGCTCATCATCA 186
140 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 199
187 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 246
200 CGGAATTTTGGAAACAGAGATGCTGTGAGTGAATTAAGGAGATTAATCATCACAAG 259
247 CGGAATTTTGGAAACAGAGATGCTGTGAGTGAATTAAGGAGATTAATCATCACAAG 306
260 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGAGATTTGAA 319
307 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGAGATTTGAA 366
320 GATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 379
367 GATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 426
380 CGGAACAGAGGCTTTTCCAGAGAGCGGCTATGAGATGGAATTAATTAAGAGCTTCA 439
427 CGGAACAGAGGCTTTTCCAGAGAGCGGCTATGAGATGGAATTAATTAAGAGCTTCA 486
440 GGGCCATACAGAGAGGTGGAAGAGTATTTCCAGAGTTGCGGTGAGAGATTTGCTCA 499
487 GGGCCATACAGAGAGGTGGAAGAGTATTTCCAGAGTTGCGGTGAGAGATTTGCTCA 546
500 GGAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 559
547 GGAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 606
560 GGTCTAGAGACCAATTAATGAGTGAACAGATTAATGATTAATTTCTCAAGAGAGT 619
607 GGTCTAGAGACCAATTAATGAGTGAACAGATTAATGATTAATTTCTCAAGAGAGT 666
620 GGCAGTGGAGTGAAC 635
667 GGCAGTGGAGTGAAC 682

RESULT 5
LOCUS B1560849 693 bp mRNA linear EST 05-SEP-2001
DEFINITION 60325401BFL NIH_MGC_97 Homo sapiens cDNA clone IMAGE:529644 5',
B1560849 mRNA sequence.
ACCESSION B1560849
VERSION B1560849.1 GI:15448163
SOURCE EST.
ORGANISM Homo sapiens (human)
Homosapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES

Email: cga@b1560849.nih.gov
Tissue Procurement: Mikhail Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHRI), Shiroki
Tohiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM11750 row: c column: 05
High quality sequence stop: 691.
Location/Qualifiers

1.693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:529644"
/lab_host="DH10B"
/clone_id="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
(Gcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.2 kb and normalized to 10⁵. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Garnicki, in
preparation). Library constructed by M. Brownstein
(NIH/NHRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN

Query Match 27.7%; Score 616; DB 12; Length 693;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATATTGAG 79
78 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATATTGAG 137
80 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCCAGCTCATCATCA 139
138 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCCAGCTCATCATCA 197
140 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 199
198 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 257
200 CGGAATTTTGGAAACAGAGATGCTGTGAGTGAATTAAGGAGATTAATTAAGAGCTTCA 259
258 CGGAATTTTGGAAACAGAGATGCTGTGAGTGAATTAAGGAGATTAATTAAGAGCTTCA 317
260 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGAGATTTGAA 319
318 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGAGATTTGAA 377
320 GATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 379
378 GATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 437
380 CGGAACAGAGGCTTTTCCAGAGAGCGGCTATGAGATGGAATTAATTAAGAGCTTCA 439
438 CGGAACAGAGGCTTTTCCAGAGAGCGGCTATGAGATGGAATTAATTAAGAGCTTCA 497
440 GGGCCATACAGAGAGGTGGAAGAGTATTTCCAGAGTTGCGGTGAGAGATTTGCTCA 499
498 GGGCCATACAGAGAGGTGGAAGAGTATTTCCAGAGTTGCGGTGAGAGATTTGCTCA 557
500 GGAATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 559
558 GGAATGATGATGATCTGCTGTTTCTGAGAGAGCTTAATATGATCTGCGAAGTAAATCCACA 617
560 GGTCTAGAGACCAATTAATGAGTGAACAGATTAATGATTAATTTCTCAAGAGAGT 619

Db 618 GCTTCTAGAACCAAGTATTAGTGCACAGTATGATGATCTTCTCAACAGCAAGT 677

QY 620 GGCAGTGGAAAGTGAAC 635

Db 678 GGCAGTGGAAAGTGAAC 693

RESULT 6
BG717110 798 bp mRNA linear EST 08-MAY-2001
LOCUS 60268919.F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4821503 5',
DEFINITION mRNA sequence.
ACCESSION BG717110
VERSION BG717110.1 GI:13996297
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM10728 row: e column: 24
High quality sequence stop: 770.
Location/Qualifiers
1..798
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4821503"
/lab_host="DH108"
/clone_lib="NIH_MGC_97"
/note="Organ: Testis; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.2 kb and normalized to 80% 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN

Query Match 27 5%; Score 612; DB 12; Length 798;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 29 TGGGAGCAGAAATCAACCTCATATGCTTCTCATATTTCCATATTTGAGAGGATAG 88
Db 66 TGGGAGCAGAAATCAACCTCATATGCTTCTCATATTTGAGAGGATAG 125
QY 89 TTTTCTGGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 148
Db 126 TATCTGGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 185
QY 149 GATGACCTTCTGAGAGATCATTTCAATGAAAGTGATTTGCTCGGGCGGAAATTT 208
Db 186 GATGACCTTCTGAGAGATCATTTCAATGAAAGTGATTTGCTCGGGCGGAAATTT 245
QY 209 GGAACAGAGATGCTGCTGATGTAATTAAGCAGATTAATATCATTCACATGGGTGTTT 268

Db 246 GAAAACAGAGATGCTGCTGATGTAATTAAGCAGATTAATATCATTCACATGGGTGTTT 305

QY 269 GGAATGGAAAGATTTTGGAAACAGAGTTTCAACAGCAGGTTTGAAGATGTGAT 328

Db 306 GGAATGGAAAGATTTTGGAAACAGAGTTTCAACAGCAGGTTTGAAGATGTGAT 365

QY 329 AGCTTGCTTCTGAGAGAGTCTAGTATATGACTCCGAATATCCAAACGGAACAGA 388

Db 366 AGCTTGCTTCTGAGAGAGTCTAGTATATGACTCCGAATATCCAAACGGAACAGA 425

QY 389 GGGTTTCCAGAGAGGCGGCTATCGAGATGGAATATTCAGAGGTTCCAGGCCATAC 446

Db 426 GGGTTTCCAGAGAGGCGGCTATCGAGATGGAATATTCAGAGGTTCCAGGCCATAC 485

QY 449 AGAAGAGTGGAAAGAGTATGTTTCCGAGGTTCCGAGAGATTTGCTTGAAGATCCA 508

Db 486 AGAAGAGTGGAAAGAGTATGTTTCCGAGGTTCCGAGAGATTTGCTTGAAGATCCA 545

QY 509 AATATGACTTGAACCCAGAGATGATGAGGCGCTGGTGGCTTTTGGTCTTGA 568

Db 546 AATATGACTTGAACCCAGAGATGATGAGGCGCTGGTGGCTTTTGGTCTTGA 605

QY 569 AGACCAATTAATGAGTGGACAGTAAATGATGATCTTCTCAACAGAGTGGCAGTGA 628

Db 606 AGACCAATTAATGAGTGGACAGTAAATGATGATCTTCTCAACAGAGTGGCAGTGA 665

QY 629 AGTAAAGAGTGGTTCAAAGGTTTAAATGAAGATTAACAGGCTTGGAAAGAT 688

Db 666 AGTAAAGAGTGGTTCAAAGGTTTAAATGAAGATTAACAGGCTTGGAAAGAT 725

QY 689 TCT 691

Db 726 TCT 728

RESULT 7
CD557160 885 bp mRNA linear EST 11-JUN-2003
LOCUS AGENCOURT 14413590 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION IMAGE:30387531 5', mRNA sequence.
ACCESSION CD557160
VERSION CD557160.1 GI:31583228
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Daniela S. Gerhard, Ph.D.
COMMENT Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDMA57 row: p column: 04
High quality sequence stop: 671.
Location/Qualifiers
1..885
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30387531"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"

FEATURES
source

[illegible]

Db	Query Match	Score	DB 13	Length	792
304	GGTGGTTTGGAGTTGGAAAGAGTTTGGAAACAGAGTTTTCGAACAGCAGGTTTGAA	24.1%	Best Local Similarity	100.0%	Pred No. 5e-274;
Qy	320 GATGTGATAGCTCTGCTGTTTCTGGAGAGAGTCTAGTAATGACTCGGAAGATATACCAACA	379			
Db	364 GATGTGTATAGCTCTGCTGTTTCTGGAGAGAGTCTAGTAATGACTCGGAAGATATACCAACA	423			
Qy	380 CGGAACAGAGGGTTTCCAGAGAGGGCGGCTATGAGATGTGAAATTAATTCAGAACTTCA	439			
Db	424 CGGAACAGAGGGTTTCCAGAGAGGGCGGCTATGAGATGTGAAATTAATTCAGAACTTCA	483			
Qy	440 GGGCCATACAGAAAGAGTGAAGAGATGATTTCCAGAGTTCCGAGGTTCCGTGAGAGATTGGTCTA	499			
Db	484 GGGCCATACAGAAAGAGTGAAGAGATGATTTCCAGAGTTCCGAGGTTCCGTGAGAGATTGGTCTA	543			
Qy	500 GGAATCCAAATTAAGACTTAGACCCAGACGAATATATGACAGCGGCACTGTGGCTTTT	559			
Db	544 GGAATCCAAATTAAGACTTANACCACAGCAATATATGACAGCGGCACTGTGGCTTTT	603			
Qy	560 GATTCTAGAAACACAGATTAATGATGGCACAGGTATGTGATCTTCTCA	609			
Db	604 GATTCTAGAAACACAGATTAATGATGGCACAGGTATGTGATCTTCTCA	653			
RESULT 10	BQ429126	792 bp	mRNA	linear	EST 24-MAY-2002
LOCUS	AGNCOURT 7901787 NIH_MGC_82 Homo sapiens	cdna	clone	IMAGE:6104663	
DEFINITION	5', mRNA sequence.				
ACCESSION	BQ429126				
VERSION	BQ429126.1	GI:21168202			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	NIH-MGC htcp://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: CLONTECH				
	cDNA Library Preparation: CLONTECH Laboratories, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)				
	DNA sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LMNL at:				
	http://image.lmnl.gov				
	Plate: LICM2342 row: n column: 24				
	High quality sequence, stop: 496.				
FEATURES	Location/Qualifiers				
source	1..792				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:6104663"				
	/lab_host="DH10B (TI phase-resistant)"				
	/clone_id="NIH_MGC_82"				
	/note="Organ: testis; Vector: pDNR-LTB (Clontech); Site 1:				
	SfiI (ggcgccctcgccg); Site 2: SfiI (ggccatcggcc); 5' and				
	3' adaptors were used in cloning as follows: 5' adaptor				
	sequence: 5'-CAGCGCATTAATGCGC-3' and 3' adaptor sequence:				
	5'-ATTCTAGAGCCCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,				
	C, or G and N = A, C, G, or T). Average insert size				
	1.35 kb (range 0.9-4.0 kb). 14/15 clones contained				
	inserts by PCR. This library was enriched for full-length				
	clones and was constructed by Clontech Laboratories (Palo				
	Alto, CA)."				

Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GATGAGATTGGGAGAGAGAAATCAACCTCATATGTCTTCCATGTCCATATTTGAG 79
Db 67 GATGAGATTGGGAGAGAGAAATCAACCTCATATGTCTTCCATGTCCATATTTGAG 126
QY 80 AAGATATGTATTTCTGGAGAAATGAGACATTTTACGAGACTCCAGCTTCATATCA 139
Db 127 AAGATATGTATTTCTGGAGAAATGAGACATTTTACGAGACTCCAGCTTCATATCA 186
QY 140 GAAATGATGATGAGACCTTCTCGAAGAGATCATTTCCATGAAAAGTGGATTGCTG 199
Db 187 GAAATGATGATGAGACCTTCTCGAAGAGATCATTTCCATGAAAAGTGGATTGCTG 246
QY 200 CGGAATTTTGGAAACAGAGATGCTGTGATGTATATAGAGATATATATCCAAAG 259
Db 247 CGGAATTTTGGAAACAGAGATGCTGTGATGTATATAGAGATATATATCCAAAG 306
QY 260 GGTGATTTTGGAGTTGGAAAGAGTTTGGAAACAGAGATTTTCCAAACAGAGTTGAA 319
Db 307 GGTGATTTTGGAGTTGGAAAGAGTTTGGAAACAGAGATTTTCCAAACAGAGTTGAA 366
QY 320 GATGTGATAGCTCTGCTTTCTGGAGAGAGTCTATGATATGACTGCGAAGTAAATCCAA 379
Db 367 GATGTGATAGCTCTGCTTTCTGGAGAGAGTCTATGATATGACTGCGAAGTAAATCCAA 426
QY 380 CGGAAGAGAGGTTTCCAAAGAGAGGCTATCGAGATGGAATATATCCAAAGCTCA 439
Db 427 CGGAAGAGAGGTTTCCAAAGAGAGGCTATCGAGATGGAATATATCCAAAGCTCA 486
QY 440 GGGCATATACAGAGAGTGGAGAGAGTATTTCCAGAGTTCCCTGGAGAGATTTGCTCA 499
Db 487 GGGCATATACAGAGAGTGGAGAGAGTATTTCCAGAGTTCCCTGGAGAGATTTGCTCA 546
QY 500 GGAAGTCCAAATATATGACTTATGACCCAGAGATGATGAGAGGCACTGGTGGCC 554
Db 547 GGAAGTCCAAATATATGACTTATGACCCAGAGATGATGAGAGGCACTGGTGGCC 601

RESULT 11

CD557137 914 bp mRNA linear EST 11-JUN-2003
LOCUS ABEHCOURT_14413272 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION IMAGE:30390297 5', mRNA sequence.
ACCESSION CD557137
VERSION CD557137.1 GI:31583205
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 914)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
CONTACT: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM465 row: c column: 10
High quality sequence start: 42
High quality sequence stop: 582.
Location/Qualifiers
1..914
/organism="Homo sapiens"

FEATURES

source
1..914
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390297"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_11b="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-Sport6.1; Site: 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
library."

ORIGIN

Query Match 23.4%; Score 521; DB 14; Length 914;
Best Local Similarity 100.0%; Pred. No. 1.6e-266; Indels 0; Gaps 0;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 TTGGAAGCCAGAAATTTCTTTTGGAGCTGTGTAAAGCTGTTGTAATATGAGGG 1218
Db 105 TTGGAAGCCAGAAATTTCTTTTGGAGCTGTGTAAAGCTGTTGTAATATGAGGG 164
QY 1219 AACCCAGCTGGGACATTTCAATTGCACAATAGTACAAAGCTGTATATATATGCTAC 1278
Db 165 AACCCAGCTGGGACATTTCAATTGCACAATAGTACAAAGCTGTATATATATGCTAC 224
QY 1279 TCTGGAAGACTGATGATATCATATAGGCAAGAAAGATTGGTCTCAACAGATCAATA 1338
Db 225 TCTGGAAGACTGATGATATCATATAGGCAAGAAAGATTGGTCTCAACAGATCAATA 284
QY 1339 CTATGTTTGGATGAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1398
Db 285 CTATGTTTGGATGAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 344
QY 1399 GTTAAATTTCTGCCCAGAAATGCCATCAAGAGAAAGAGCCAAACCTTATGTTCA 1458
Db 345 GTTAAATTTCTGCCCAGAAATGCCATCAAGAGAAAGAGCCAAACCTTATGTTCA 404
QY 1459 AACTTTTCCAGAGAAATTCAAAGGTGGCTGACAGATTTTAAAGTCAAAATATCTG 1518
Db 405 AACTTTTCCAGAGAAATTCAAAGGTGGCTGACAGATTTTAAAGTCAAAATATCTG 464
QY 1519 TCTTGTCTGTGGACAAAGTGGTGGAGCATGTAAGATGTTTACGAGACCTTCCAA 1578
Db 465 TCTTGTCTGTGGACAAAGTGGTGGAGCATGTAAGATGTTTACGAGACCTTCCAA 524
QY 1579 TGGCCAGTTCTCAAAAAGAGAAAGAGCTGTTGAAATTTCTGCGAAACATAGGGATGA 1638
Db 525 TGGCCAGTTCTCAAAAAGAGAAAGAGCTGTTGAAATTTCTGCGAAACATAGGGATGA 584
QY 1639 AACTATGCTCTTTGTTGAAATCTAAGAAAAAGACAGATTTTA 1679
Db 585 AACTATGCTCTTTGTTGAAATCTAAGAAAAAGACAGATTTTA 625

RESULT 12

BTG19449 747 bp mRNA linear EST 08-MAY-2001
LOCUS BTG19449
DEFINITION 602690172F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822528 5',
mRNA sequence.
ACCESSION BTG19449
VERSION BTG19449.1 GI:13998636
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 747)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

Oy		260	GGTGGTTTGGAGTGGAAAGGTTTTGGAAAACAGAGGTTTTTCAAACAGCAGGTTTGA	319
Dd		297	GGTGTTTGGAGTTGGAAAAGGTTTTGGAAAACAGAGGTTTTTCAAACAGCAGGTTTGA	356
Oy		320	GATGCTGTAGCTCTGTTTCTGGAGAAGTCCTAATAATGACTCGGAGATPATCCACA	379
Dd		357	GATGCTGTATACCTCTGGTTTCTGGAGAAGTCCTAATAATGACTCGGAGATPATCCACA	416
Oy		380	CGAAACACAGAGGTTTTCCAGAGAGAGCGGCTATGCAGATGGAATAATCA- GAAGCTTC	438
Dd		417	CGAAACACAGAGGTTTTCCAGAGAGAGCGGCTATGCAGATGGAATAATCA- GAAGCTTC	476
Oy		439	AGGCCCATACAGAAAGAGGTGGAAAGGTAGTTTCCGAGGTTGCCGTGAGGAATTGGTCT	498
Dd		477	AGGCCCATACAGAAAGAGGTGGAAAGGTAGTTTCCGAGGTTGCCGTGAGGAATTGGTCT	536
Oy		499	AGGAAGTCCAAATATATGACTTAGACCAGACGAAATGTATGACAGGCACTGGTGCCCTTT	558
Dd		537	AGGAAGTCCAAATATATGACTTAGACCAGACGAAATGTATGACAGGCACTGGTGCCCTTT	596
Oy		559	TGGTTCTGAAAGAACCGATTAATAGTGGCACAGTATATGTTACTTCTCAAAGCAGAAG	618
Dd		597	TGGTTCTGAAAGAACCGATTAATAGTGGCACAGTATATGTTACTTCTCAAAGCAGAAG	656
Oy		619 T 619		
Dd		657 T 657		
RESULT 14				
LOCUS	CDS57866			
DEFINITION	AGENCOURT 14423254 NIH MGC 180 Homo sapiens cDNA clone	933 bp	mRNA	linear EST 11-JUN-2002
ACCESSION	CDS57866			
VERSION	CDS57866.1	GI:31583934		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	NIH-MGC http://mgi.nci.nih.gov/.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999) Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rml0A07 Bethesda, MD 20892 Email: cgabds@mail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: NDAM464 Row: g Column: 21 High quality sequence start: 19 High quality sequence stop: 592.			

FEATURES
source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390020"
/lab_host="DH10B-Txn A ( T1 and T5 phage resistance)"
/clone_lib="NIH_MGC_189"
/name="Orogen: Testis; Vector: pCMV-Sport6.1; Site.1: NotI
site.2: EcoRV (destroyed); Library is ciso-df primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC

```

ORIGIN	Library."
Query Match	21.2%; Score 471; DB 14; Length 933;
Best Local Similarity	100.0%; Pred. No. 8.ee-240;
Matches 471; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY	1012	AACAGGGCTTGGGAAGACTGCGCGCTTTTCTCTCCACAAATTTTGGCTCATATGATGCATGA	107
Db	51	AACAGGGCTTGGGAAGACTGCGCGCTTTTCTCTCCAAATTTTGGCTCATATGATGCATGA	110
QY	1072	TGGAATACTCCAGTCGTTTAAAGATGACAGAACCCAGAGTATATATGTATGCACC	113
Db	111	TGGAATACTCCAGTCGTTTAAAGATGACAGAACCCAGAGTATATATGTATGCACC	170
QY	1132	AACTGGAAATTGGTCAACAGATTTATTTGGAAAGCCAGAAAATTTTCTTTTGGCACTTG	119
Db	171	AACTGGAAATTGGTCAACAGATTTATTTGGAAAGCCAGAAAATTTTCTTTTGGCACTTG	230
QY	1192	TGTAAGAGCTTGTGTATATATGAGGGGAACCCAGCTGGGACATTCGACAAATAGT	125
Db	231	TGTAAGAGCTTGTGTATATATGAGGGGAACCCAGCTGGGACATTCGACAAATAGT	290
QY	1252	ACAAGGCTGTATATATTTATGTGCTACTCTCTGGAAAGCTGATGATATCATATGSCAAGA	131
Db	291	ACAAGGCTGTATATTTATGTGCTACTCTCTGGAAAGCTGATGATATCATATGSCAAGA	350
QY	1312	AAAGATTTGTCGTCAAACAGATCAAAATCTTAGTTTGGATGAGGTGATCGCATGTGGA	137
Db	351	AAAGATTTGTCGTCAAACAGATCAAAATCTTAGTTTGGATGAGGTGATCGCATGTGGA	410
QY	1372	TATGGGTTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAATGCCATCAAGA	143
Db	411	TATGGGTTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAATGCCATCAAGA	470
QY	1432	ACAGGCGCAACCCCTTATGTCACAGTGCACCTTTTCCAGAGAAATTCGAAG	148
Db	471	ACAGGCGCAACCCCTTATGTCACAGTGCACCTTTTCCAGAGAAATTCGAAG	521

[illegible]

Email: cgaapds-remail.nih.gov
 Tissue Procurement: M.Klos Falkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: KGC Clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM10729 row: p column: 22
 High quality sequence stop: 651.
 Location/Qualifiers
 1..651
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

```
/clone="IMAGE:4822149"  
/lab host="DH10B"  
/clone lib="NIH_MGC_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI  
(GCGAG); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
insert size 2.2 kb and normalized to R01 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIMH/NHGRI, National Institutes of Health). Note: this is  
a NIH_MGC Library."
```

ORIGIN

Query Match

21.1%; Score 469; DB 12; Length 651;

Best Local Similarity 100.0%; Pred. No. 9,6e-239;

Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 29 TGGGAAGCAGAAATCAACCTCATATGCTCTCTATGTTCCCATATTTGAGAAAGATAG 88  
DB 77 TGGGAAGCAGAAATCAACCTCATATGCTCTCTATGTTCCCATATTTGAGAAAGATAG 136  
QY 89 TATTTGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 148  
DB 137 TATTTGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 196  
QY 149 GATGACCTTCTGAAAGAGATCATTTTCATGAAAGTGATTTGCTCTGCGCGAAATTT 208  
DB 197 GATGACCTTCTGAAAGAGATCATTTTCATGAAAGTGATTTGCTCTGCGCGAAATTT 256  
QY 209 GGAACAGAGATGCTGAGTGTATTAAGCGAGATTAATACCAATGGATGGTGT 268  
DB 257 GGAACAGAGATGCTGAGTGTATTAAGCGAGATTAATACCAATGGATGGTGT 316  
QY 269 GGAAGTGAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGTTTGAAGATGAT 328  
DB 317 GGAAGTGAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGTTTGAAGATGAT 376  
QY 329 AGCTCTGCTTCTGAGAGAGCTAGTAATGATGGAAGATTAATCAACAGGAAACAG 388  
DB 377 AGCTCTGCTTCTGAGAGAGCTAGTAATGATGGAAGATTAATCAACAGGAAACAG 436  
QY 389 GGGTTTCCAAAGAGCGGCTATCGAGATGGAATTAATTCAGAACTTCAGGCGCATAC 448  
DB 437 GGGTTTCCAAAGAGCGGCTATCGAGATGGAATTAATTCAGAACTTCAGGCGCATAC 496  
QY 449 AGAAGAGGTGAAGAGATGTTCCAGAGTTGCCGTGAGAGATTTGTC 497  
DB 497 AGAAGAGGTGAAGAGATGTTCCAGAGTTGCCGTGAGAGATTTGTC 545
```

Search completed: July 27, 2004, 18:08:12

Job time : 5702 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 03:14:19 ; Search time 5696 Seconds
(without alignments)
11659.672 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224
Sequence: 1 actgtgaagtcacatgsggg999.....aagctctgtgtttgatgca 2224

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estlm:*
4: em_estlm:*
5: em_estlm:*
6: em_estlm:*
7: em_estlm:*
8: em_estlm:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_hic:*
13: gb_hic:*
14: gb_hic:*
15: em_estlm:*
16: em_estlm:*
17: em_gss_hum:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_hum:*
21: em_gss_hum:*
22: em_gss_hum:*
23: em_gss_hum:*
24: em_gss_hum:*
25: em_gss_hum:*
26: em_gss_hum:*
27: em_gss_hum:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2200.8	99.0	2523	11	BC030638 Homo sapi
2	1539.8	63.2	2753	11	AK014844 Mus muscu
3	772.2	34.7	896	14	CD557531 AGENCOURT
4	741.8	33.4	850	14	CD251828 AGENCOURT

5	733.8	33.0	933	14	CD557866
6	729.6	32.8	885	14	CD557160 AGENCOURT
7	694.4	31.2	798	12	BG717110 B1464267
8	652.2	29.3	822	12	B1464267
9	649	28.2	914	14	CD557137
10	643.2	28.9	792	13	BQ428126
11	636.2	28.6	747	12	BG719449
12	631.8	28.4	682	12	BG718234
13	631.8	28.4	693	12	B1560849
14	631.2	28.4	868	14	CD558073
15	627	28.2	830	12	B1462207
16	623.8	28.0	918	12	B1561001
17	621.8	28.0	846	12	BG718634
18	617.6	27.8	981	14	CD359812
19	608.8	27.4	699	12	BG717822
20	608	27.3	758	12	B1559479
21	607.8	27.3	939	14	CD300339
22	606.2	27.3	861	12	B1461393
23	604.8	27.2	657	12	BG717812
24	596	26.8	743	12	B1463502
25	593.2	26.7	884	12	B1560499
26	591	26.6	823	13	B0939339
27	577.4	26.0	805	14	CB230120
28	575.8	25.9	651	12	BG717354
29	571.2	25.7	700	12	BG718861
30	555.8	25.0	847	12	B1562624
31	554	24.9	810	12	B1463846
32	536.2	24.1	841	12	B1459115
33	515.4	23.2	917	12	B1463447
34	483.6	21.7	546	12	BMS39751
35	482.6	21.7	561	14	CD561008
36	469.4	21.1	690	14	CD767258
37	449.8	20.2	670	13	BQ443738
38	445.2	20.0	482	9	AL042306
39	420.2	18.9	780	10	AW153705
40	405.8	18.2	759	14	CK026554
41	393.2	17.7	635	13	BY716292
42	389	17.6	512	28	CC183462
43	389	17.5	726	10	BF611759
44	380.2	17.1	604	12	B1460628
45	371	16.7	4617	11	BC007668

ALIGNMENTS

RESULT 1
BC030638
LOCUS
DEFINITION
Homo sapiens cDNA clone IMAGE:4822528, containing frame-shift errors
ACCESSION
BC030638
VERSION
BC030638.1
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 2523)
Srausderg,R.L., Feingold,R.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buelow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Udell,T.B., Toshynski,S., Carninci,P., Prange,C., Kane,S.S., Loguercio,N.A., Peters,G.J., Abramson,R.D., Muliyil,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Hale,S., Garcia,A.M., Gay,L.J., Huijck,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Rahy,J., Helton,E., Ketterman,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,W., Madan,A., Young,A.C., Shevchenko,Y.,

Journal Title
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butler, F.D., Krzywicki, M.I., Skalska, U., Smalusz, D.E.,
Scherer, A., Schein, J.E., Jones, S.D., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2 (bases 1 to 2523)
Strausberg, R.
Direct Submission
Submitted (20-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 33 Row: d Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9507236
This clone has the following problem: frame shifted.
Location/Qualifiers
1..2523
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4822528"
/issue_type="Testis"
/issue_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

FEATURES

source

ORIGIN

Query Match 99.0%; Score 2200.8; DB 11; Length 2523;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2216; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 ACTGAGTCACCATGGGGATGAGATTGGAGAGCAAAATCAACCTCATATGCTTC 60
DB 85 ACTTGAGCCACCATGGAGATGAGATTGGAGAGCAAAATCAACCTCATATGCTTC 144
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATTCTGGAGAAATGAGACAATTTTAAAG 120
DB 145 CTATGTTCCCATATTTGAGAGAGATAGTATTCTGGAGAAATGAGACAATTTTAAAG 204
QY 121 GACTCCAGCTTCATCTCAAAATGAGATGAGACCTTCTGAGAGATCATTTTCAGAA 180
DB 205 GACTCCAGCTTCATCTCAAAATGAGATGAGACCTTCTGAGAGATCATTTTCAGAA 264
QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGGTGAAGTAATAAGCG 240
DB 265 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGGTGAAGTAATAAGCG 324
QY 241 AATATATACATCAACATGGTGTGTTTGGAGTTGAAAAGTTTGGAAAAGAGGTTT 300
DB 325 AATATATACATCAACATGGTGTGTTTGGAGTTGAAAAGTTTGGAAAAGAGGTTT 384
QY 301 TTCAAACAGCAGTTTGAAGATGATAGTCTGCTGTTTGGAGAGAGTCTAGTAATGA 360
DB 385 TTCAAACAGCAGTTTGAAGATGATAGTCTGCTGTTTGGAGAGAGTCTAGTAATGA 444

QY 361 CTGCGAAGATATATCCAAACAGGAAACAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 420
DB 445 CTGCGAAGATATATCCAAACAGGAAACAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 504
QY 421 AATATATCAAGAGCTTGAAGGCAATCAGAGAGGTTGAAAGAGTGTGTTCCAGGTTG 480
DB 505 AATATATCAAGAGCTTGAAGGCAATCAGAGAGGTTGAAAGAGTGTGTTCCAGGTTG 564
QY 481 CCGTGAGAGATTTGGTCTAGGAAGTCCAAATATGACTTGAACCCAGACGATATATGA 540
DB 565 CCGTGAGAGATTTGGTCTAGGAAGTCCAAATATGACTTGAACCCAGACGATATATGA 624
QY 541 GCGACCTGGTGGCTTTTGGTTCTTGAAGCCGATTTTAAGTGGACAGGTATGTGTA 600
DB 625 GCGACCTGGTGGCTTTTGGTTCTTGAAGCCGATTTTAAGTGGACAGGTATGTGTA 684
QY 601 TACTTCTCAAAGCAGAAAGTGGCAGTGAAGTGAACGAGTGTGTTCAAGGTTTAAATGA 660
DB 685 TACTTCTCAAAGCAGAAAGTGGCAGTGAAGTGAACGAGTGTGTTCAAGGTTTAAATGA 744
QY 661 AGAAGTATTAACAGGCTCTGGAAAGAAATTTCTTGAAGTCAAGACAGAGAGAGAAAG 720
DB 745 AGAAGTATTAACAGGCTCTGGAAAGAAATTTCTTGAAGTCAAGACAGAGAGAGAAAG 804
QY 721 TAGTGATATCAAGACCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATAGGA 780
DB 805 TAGTGATATCAAGACCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATAGGA 864
QY 781 CTCCATCTTTGCACTTATGACAGAGCATTAACCTTGAACAAATACGACATATTTCTGT 840
DB 865 CTCCATCTTTGCACTTATGACAGAGCATTAACCTTGAACAAATACGACATATTTCTGT 924
QY 841 GGAAGTCTGGAATGATGACACAGAGAAATTTGACTTGGAGAAAGCTATATCTG 900
DB 925 GGAAGTCTGGAATGATGACACAGAGAAATTTGACTTGGAGAAAGCTATATCTG 984
QY 901 TCAGACATGAATTAACACATTTGCTAAAGTGTATTAATTAATTAATTAATTAATTA 960
DB 985 TCAGACATGAATTAACACATTTGCTAAAGTGTATTAATTAATTAATTAATTAATTA 1044
QY 961 ATACAGATTTCTATCTATCTTGAAGAGAGATTTGATGCTTGTGCTCAACAGGCTC 1020
DB 1045 ATACAGATTTCTATCTATCTTGAAGAGAGATTTGATGCTTGTGCTCAACAGGCTC 1104
QY 1021 TGGAGAGACTGGGCTTTTCTCCTCAATTTTGGCTCATATGATGATGATGATGATGAT 1080
DB 1105 TGGAGAGACTGGGCTTTTCTCCTCAATTTTGGCTCATATGATGATGATGATGATGAT 1164
QY 1081 TGCAGTCTGTTTAAAGAGTTGCAAGAACCAAGTATTTATTTGACCACTGAGAA 1140
DB 1165 TGCAGTCTGTTTAAAGAGTTGCAAGAACCAAGTATTTATTTGACCACTGAGAA 1224
QY 1141 ATTGATCAACAGATTTATTTGAGAGCCAGAAATTTCTTTGGAGCTTGTGTAAAGC 1200
DB 1225 ATTGATCAACAGATTTATTTGAGAGCCAGAAATTTCTTTGGAGCTTGTGTAAAGC 1284
QY 1201 TGTGTTATATATGAGGAGAAACCAAGTGGACATTCATTTGCAAAATATGTAAGGCTG 1260
DB 1285 TGTGTTATATATGAGGAGAAACCAAGTGGACATTCATTTGCAAAATATGTAAGGCTG 1344
QY 1261 TAAATATATATGAGTATCTGAGAGATGATGATATCATATGAGAAAGAAAGTGTG 1320
DB 1345 TAAATATATATGAGTATCTGAGAGATGATGATATCATATGAGAAAGAAAGTGTG 1403
QY 1321 TCTCAAACAGATCAAAATCTTATTTTGAAGAGTGAATGCAATGATGATGATGATG 1380
DB 1404 TCTCAAACAGATCAAAATCTTATTTTGAAGAGTGAATGCAATGATGATGATGATGATG 1463
QY 1381 TGGTCCAGAAATGAAAGATTAATTTCTTGGCCAGAAATGCAATCAAGAGAAACAGGCTCA 1440
DB 1464 TGGTCCAGAAATGAAAGATTAATTTCTTGGCCAGAAATGCAATCAAGAGAAACAGGCTCA 1523

FEATURES
SOURCE

Location/Qualifiers
1..2759
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4921508G22"
/db_xref="MGI:1910815"
/db_xref="taxon:10090"
/clone="4921508G22"
/sex="male"
/cissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
42..2150
/note="unannotated protein product; DEAD (aspartate-glutamate-alanine-aspartate) box polypeptide 4 (MGI:MGI:102670)
CDs
putative
/codon_start=1
/protein_id="BAB29578.1"
/db_xref="GI:12852922"
/translation="MGDEDEWEAEILKPHVSSYPVPEKDKYSSGANGDTFNTSASD
IGSSKKENTSTTGGRGKRGKNGRPLNNKPEEGSSGPKMESNNDEEDNTRRGF
SKRGCGODNDEBASPPRRGRGSGFRGGRGFGDGRGSESDDGDTORGGLGSR
KPAASDGNEDTIOSSRSGRGKGLNEVTVGSKNWKSETEGESDSQGRVT
YLPPEDEDSIFAHYQIGINEFDKYDTILIVSGHDAPALITFEANLCOTLNNI
AKAGYKTLIPVKYSIPVLGRDLMACTGSGKTAFLILAHMRDGTIAIRFK
ELQEPICIVAPRELINOIYLEAKFSFGTCRAVAVIGTQFGHSYROIYOGNII
CATPGLMDIGEKIKGVKXIVLDEADMLDMGREGEMKILISCGMPKRGROT
LFSATPEIOLAGDPLKSSYLPYAVONGACRDNVQITLQYQISKEKLEIYL
RNIGDERIVFYETKKADPIATFLQEKISTISHDBQRERQALSPFTDSDNH
LVATSVARGLDLENVQIVINPDLSTIDYVHRIGRTGRCNTGRLAISFTDSDNH
LAQPLKVLSDAQDVPALWLEIAFSTYVPSFSSSTRGAVFASVDTRKMYQGHLL
NTAGISSQAPNVDDESMD"
2734..2739
polya_signal
/note="putative"
2759
polya_site
/note="putative"
ORIGIN

Query Match 69.2%; Score 1539.8; DB 11; Length 2759;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 262; Indels 23; Gaps 5;

QY 181 AAGTGAATTTGCTCTGGGCGGAATTTGAAAACAGAGTGTGTGATGTAATGACG 240
DB 134 AATGAGACACTTTTACAGACCTTCACTTCACTGATATTGGCGAGTCTAATAAAA 193
QY 241 AGATAATACATCCACATGGTGTGTTTGAAGTGAAGAAGTTTGGAAAACAGAGTT 300
DB 194 AGGAACACATCTACACATGTGTGCTTTGGAAGAGAAAAGGCTTTGAAAACAGAGTTT 253
QY 301 TTCAAACAGCAGTTTGAAGATGTGATAGCTCTGCTTTCTGAGAGAGTCTAATGTA 360
DB 254 TTAAATATCAAGTTTGAAGAGTGTAGCTCTGCTTTCTGGAAGAGTCTAATATGA 313
QY 361 CTGCGAAGATATCCACAACGGAACAGAGGTTTCCAAAGAGCGGCTATCCAGATG 420
DB 314 CTGTGAAGATATCAAGCTCGAAGCAGAGGTTTTCACAGGAGTGTGCTGCAAGATG 373
QY 421 AATAATTCAGAGCTTCAGGCGCATACAGAAAGGTGGAAGAGGTAGTTTCCAGAGTTG 480
DB 374 AAACGATTCAGAGCATCAGGCCCGTTTCAAGAGAGGGAAGAGCACTTTCCAGAGCTG 433
QY 481 CCGTGAAGATTTGTGTTAGAAAGTCCAAATATGACTTGAACCCAGACGATGTATGCA 540
DB 434 CCGTGAAGATTTGTGTTAGAAAGCAAAATAGTAAATCTGACCAAGATCAGGGAGACA 493
QY 541 GCGCACTGGTGGCTTTTGTGTTCTAGAAAGCAGATATTAAGTGCACAGGTAATGTGA 600
DB 494 GCGTGGTGGTGGCTTTTGTGTTCTAGAAAGCAGACAGCAAGTATTAAGGCAATGTGTA 553
QY 601 TACTTCTCAAGCAGAGTGGCAGTGAAGTGAACAGAGTGTGTTCAAAAGGTTTAAATGA 660

DB 554 CACTTACCAAGCAGAG-----AGTGAAGGTGTGAGGTGTTACAAAGGTTTAAATGA 607
QY 661 AGAATATTAACAGGCTCTGGAAAGAAATTTCTGGAAGTCAAGAACAGAGGAGAGAAG 720
DB 608 AGAAGTATTAACAGGTTCTGGAAGAAATTTCTGGAAGTCAAGAACAGAGGAGAGAAG 667
QY 721 TAGTATATCTCAAGACCAAAAGTACATATCCCCCTCTCTCACTTGAAGATGAGGA 780
DB 668 CAGTATATGTCAAGGTCCAAAAGTACATATATCCCCCTCTCTCAAGAGGATGAGGA 727
QY 781 CTCCATCTTTCACATTATCAGACAGCATTAATCTGCAAAATACACATATTTCTGT 840
DB 728 CTCCATCTTTCACATTATCAGACAGCATTAATCTTGAATATACACATATTTCTGT 787
QY 841 CGAAGTGTCTGGAATGATGACCAACAGCATTTCTGACTTTGAGAAGCTAATCTGTG 900
DB 788 TGAAGTATCTGACATATGACACACCGGCAATTTTCACTTTGAAAGATATCTCTG 847
QY 901 TCAGACCTGAATTAACAACATTTGCTAAAGCTGTTACTAAGTTTCTCTGTCGCAAA 960
DB 848 TCAGACCTGAATTAACAACATTTGCTAAAGCTGTTACTAAGTTTCTCTGTCGCAAA 907
QY 961 ATACAGTATTTCTATCATCTTTCAGAGACAGATTTGATGCTTGTCTCAACAGGCTC 1020
DB 908 GTACAGCATTTCCCATTTGATTAAGACAGAGATTTGATGCTTGTCTCAACAGGCTC 967
QY 1021 TGGGAAGACGCGGCTTTCTCTCAACATTTTGGCTCATATGATGATGATGGAATAC 1080
DB 968 TGGGAAGACGCGGCTTTCTCTCAACATTTTGGCTCATATGATGATGATGGAATAC 1027
QY 1081 TGCCAGTGTGTTTAAAGATGAGAGAACAGAGTATATTTGTAGCACTGACAGCA 1140
DB 1028 TGCCAGTGTGTTTAAAGATGAGAGAACAGAGTATATTTGTAGCACTGACAGCA 1087
QY 1141 ATTGCTACACAGATTTATTTGAGAGCAGAAAATTTCTTTTGGGACTTGTGTAAAGC 1200
DB 1088 ACTGATACACCAATTTACTTGAAGCAGAAAATTTCTTTTGGGACTTGTGTAAAGC 1147
QY 1201 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAAGGCTG 1260
DB 1148 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAAGGCTG 1207
QY 1261 TAATATATATGCTACTCTCTGAAAGCTGATGATATTCATATGCAAGAAAAGATG 1320
DB 1208 TAATATATATGCTACTCTCTGAAAGCTGATGATATTCATATGCAAGAAAAGATG 1267
QY 1321 TCTCAACACATCAAAATCTTATGTTTGAAGAGTGAAGTGCATGATGATGATGATG 1380
DB 1268 CTTCAACACATCAAAATCTTATGTTTGAAGAGTGAAGTGCATGATGATGATGATG 1327
QY 1381 TGGTCCAGAAATGAAGAGTTAATTTCTGCCCAGAGATGCCATCAAGAGAACAGGCCA 1440
DB 1328 TGAACACAGATGAAGAACTAATTTCTGTCCAGAGATGCCATCAAGAGAACAGGCCA 1387
QY 1441 AACCTTATGTTAGAGTGAACCTTTTCCAGAGAAATTCAAAGTTGGCTGCAGAGTTT 1500
DB 1388 AACCTTATGTTAGAGTGAACCTTTTCCAGAGAAATTCAAAGTTGGCTGCAGAGTTT 1447
QY 1501 AAAGTCAATATCTGTTTGTGCTGTTGGAACAAGTGGGTGAGCATGTAGAGATTTCA 1560
DB 1448 AAAGTCAATATCTGTTTGTGCTGTTGGAACAAGTGGGTGAGCATGTAGAGATTTCA 1507
QY 1561 GCAAGCCGTTCTCAAGCTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTG 1620
DB 1508 GCAAGCCGTTCTCAAGCTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTG 1567
QY 1621 AAACATAGGAGATGAAGAACTATGCTCTTGTGTAACCTAAGAAAAGAGAGATTTTC 1680
DB 1568 AAACATAGGAGATGAAGAACTATGCTCTTGTGTAACCTAAGAAAAGAGAGATTTTC 1627
QY 1681 TGAACCTTTCTTGTCAAGAAAATATCAACTACAGATATCAATGATGAGGAGACA 1740

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	1628	TGCAACTTTTCTTTGTGTCAGAAAAAATAATCAACTAACAAGTATTCATGGTGTATCGGAGCA	1681					
Qy	1741	GAGAGAGCGGAGCAAGCTCTTTGAGAGATTTTCGCTTTGGAAAAGTCCCAATTTCTGTGTC	1800					
Db	1688	GAGGAGAGAGAGCAAGCTCTTTGAGAGATTTTCGCTTTGGAAAAGTCCCAATTTCTGTGTC	1747					
Qy	1801	TACTTCAGTAGCTCCAGAGGGCTGATATTTGAAAATGTGCACATGTTATCAATTTTGA	1868					
Db	1748	TACTTCAGTAGCTCCAGAGGGCTGATATTTGAAAATGTGCACATGTTATCAATTTTGA	1807					
Qy	1861	TCTTCCTTTCAACATGATGTAATATGTTCAATCGAATTGGGGTACTGGTGGTGTGGGA	1920					
Db	1808	CTTTCCTTTCAACATGATGTAATATGTTCAATCGAATTGGGGTACTGGTGGTGTGGGA	1867					
Qy	1921	TACTGGCAGAGCAATTTCCCTTTTGTGATCTTGAATCGAATACCATTTTGGACAGCCCTCT	1980					
Db	1868	TACTGGCAGAGCAATTTCCCTTTTGTGATCTTGAATACCATTTTGGACAGCCCTCT	1927					
Qy	1981	AGTAAAGATATTCAGACAGATGCTCAACAGAGATTTCCGACATGTTGAGAGAAATTCCTT	2040					
Db	1928	AGTAAAGATCTGTCAAGAGCTCAACAGAGATTTCCGACATGTTGAGAGAAATTCCTT	1987					
Qy	2041	TAGTACATACAT---TCCTGGCTTCAGTGTATGACAGA---GGAAACGTGTGTGCATC	2094					
Db	1988	CAGTACCTATGTGCTCTCCAGCTTCAGTAGAGCACAAGAGGGGGGGCTGTGTTCGATC	2047					
Qy	2095	AGTTGATATCCAGAA-----AGGGCAAGACATTTGAACACAGCTGGGTTTTCTTC	2145					
Db	2048	TGTTGACACGAGAGAAATTAACAGGGCAACACAGTTGAATATAGGGGGATTTCTTC	2107					
Qy	2146	TTTCAGAGCTCCCAATCCAGTAGATGATGATGATCGATGGATTAAAGCCAA--AACATCCTT	2203					
Db	2108	TTTCAGAGCTCCCAATCCAGTAGATGATGATGATCGATGGATTAAAGCCAA--AACATCCTT	2167					
Qy	2204	CAAGTCTGTGCTTTT 2218						
Db	2168	CAAGTCTGTGATGCTTT 2182						
RESULT 3	CD557531	896 bp	mRNA	linear	EST	11-JUN-2003		
LOCUS	AGENCOURT_14413281	NIH MGC_180	Homo sapiens	cdna	clone			
DEFINITION	IMAGE:30380465 5', mRNA sequence.							
ACCESSION	CD557531							
VERSION	CD557531.1	GI:31583599						
KEYWORDS	EST.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
TITLE	NIH-MGC http://mgs.nci.nih.gov/.							
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)							
COMMENT	Unpublished (1999)							
	Contact: Daniela S. Gerhard, Ph.D.							
	Office of Cancer Genomics							
	National Cancer Institute / NIH							
	Bldg. 31 Rm10A07 Bethesda, MD 20892							
	Email: cgsabds-remail.nih.gov							
	Tissue Procurement: Dr. Michael Brownstein							
	cDNA Library Preparation: Invitrogen Corp							
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)							
	DNA Sequencing by: Agencourt Bioscience Corporation							
	Clone distribution: MGC clone distribution information can be							
	found through the I.M.A.G.E. Consortium/LNL at:							
	http://image.llnl.gov							
	Plate: NDAM465 row: 5 column: 10							
	High quality sequence stop: 592.							
FEATURES	Location/Qualifiers							
source	1..896							

Query Match	34.7%	Score 772.2	DB 14	Length 896
Best Local Similarity	97.7%	Pred. No. 3.3e-193		
Matches 794	Conservative 0	Mismatches 18	Indels 1	Gaps 1
Oy	919	CATTGCTAAAGCTGGTATATCTAAAGCTTACCTCCTGTCGCAAAATATACAGTATTTCCATCAT	978	
Db	1	CATTGCTAAAGCTGGTATATCTAAAGCTTACCTCCTGTCGCAAAATATACAGTATTTCCATCAT	60	
Oy	979	ACTTGCAGACGACGATTTGATGCGTGTGTGCTCAAAACAGGGCTGAGGAAGATCGGGGCTTT	1033	
Db	61	ACTTGCAGACGACGATTTGATGCGTGTGTGCTCAAAACAGGGCTGAGGAAGATCGGGGCTTT	120	
Oy	1039	TCTCCTACCAATTTTGGCTCATATGATGACATGATGATGATTAATCTGCCAGTCGTTTAAAGA	1098	
Db	121	TCTCCTACCAATTTTGGCTCATATGATGACATGATGATGATTAATCTGCCAGTCGTTTAAAGA	180	
Oy	1099	GTTTGCAGGAACGACAGTGTATTTATTGACACCACTCGAATTTGGTTCACACGATTTA	1155	
Db	181	GTTTGCAGGAACGACAGTGTATTTATTGACACCACTCGAATTTGGTTCACACGATTTA	240	
Oy	1159	TTTGAAGCCGAGAAATTTTCTTTTGGGACTTGTGAAGAGCTGTGTATATATATGAGGGG	1218	
Db	241	TTTGAAGCCGAGAAATTTTCTTTTGGGACTTGTGAAGAGCTGTGTATATATATGAGGGG	300	
Oy	1219	AACTCAGCTGGGACATTTCAATTGCAACAAATATGTAACAAGGCTGTATATATATATGTCATC	1277	
Db	301	AACTCAGCTGGGACATTTCAATTGCAACAAATATGTAACAAGGCTGTATATATATATGTCATC	360	
Oy	1279	TCTCGAAGAAGCTGATGATATCATAGGCAAAAGAAAGATTTGGTCTCAAAACAGATCAATA	1338	
Db	361	TCTCGAAGAAGCTGATGATATCATAGGCAAAAGAAAGATTTGGTCTCAAAACAGATCAATA	420	
Oy	1339	CTTAGTTTGGATGAACTGATGCCATGTTGGATATATGGGTTTGGTCCAGAAATGCAAGA	1399	
Db	421	CTTAGTTTGGATGAACTGATGCCATGTTGGATATATGGGTTTGGTCCAGAAATGCAAGA	480	
Oy	1399	GTTAATTTCTTCCACAGGAATGCCATCAAAAGAACAGGCCCAAAACCTTATGTTCAAGTGC	1455	
Db	481	GTTAATTTCTTCCACAGGAATGCCATCAAAAGAACAGGCCCAAAACCTTATGTTCAAGTGC	540	
Oy	1459	AACTTTTCCAGAGGAATTCAAAGGTTGGCTGACAGCTTTTAAAGTCAATTTATCTGTT	1511	
Db	541	AACTTTTCCAGAGGAATTCAAAGGTTGGCTGACAGCTTTTAAAGTCAATTTATCTGTT	600	
Oy	1519	TGTTGCTGTTGGAAGAAGTGGGTGAGAGCATGTAGAGATGTTTCAGCAGACCGTTCCTCAAGT	1578	
Db	601	TGTTGCTGTTGGAAGAAGTGGGTGAGAGCATGTAGAGATGTTTCAGCAGACCGTTCCTCAAGT	660	
Oy	1579	TGGGCAAGTTCCAAAAAGAGAAAGCTGCTGGAATTTCTGGGAAACATAGGGGATGGAAG	1633	
Db	661	TGGGCAAGTTCCAAAAAGAGAAAGCTGCTGGAATTTCTGGGAAACATAGGGGATGGAAG	720	
Oy	1639	AACATAT-GGTCTTTGTTGGAAGCTAGAGAAAGAGAGATTTTATCTGCAACTTTTCTTGTGTC	1697	
Db	721	AACATATGGGTCTTTGTTGGAAGCTAGAGAAAGAGAGATTTTATGTCACCTTTTCTTGTGT	780	
Oy	1698	AAGGAAAAATTCACATTCACAGTATCCATGGTG 1730		
Db	781	CAGAAAAATATTCACATTCACAGTATTCAGGGG 813		

RESULT 4
CD251828 850 bp mRNA linear EST 22-MAY-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITL
JOURNAL
COMMENT

CD251828 850 bp mRNA linear EST 22-MAY-2003
AGENCOURT_14206017 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30383198 5', mRNA sequence.
CD251828
CD251828.1 GI:31012294
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 850)
NIH-MGC http://imgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINTL at:
http://image.llnl.gov
Plate: NDAM46 row: k column: 15
High quality sequence stop: 626.
Location/Qualifiers
1. 850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30383198"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances) "
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

ORIGIN

Query Match 33.4%; Score 741.8; DB 14; Length 850;
Best Local Similarity 98.1%; Pred. No. 3.7e-185;
Matches 782; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

QY 10 CACCATGGGGGATGGAAGATTGGAGACGAAATCAACCTCATATGCTCTTCTATGTTCC 69
Db 7 CCCCATGGGAGAGGATGGGAGACGAAATCAACCTCATATGCTCTTCTATGTTCC 66
QY 70 CATATTGAGAGAGATAGGATCTCTGAGAGAAATGAGACATTTTAAACAGAGCTCCAGC 129
Db 67 CATATTGAGAGAGATAGGATCTCTGAGAGAAATGAGACATTTTAAACAGAGCTCCAGC 126
QY 130 TTTCATCATGAAATGATGATGACCTTCTGAGAGATCATTTTCATGAAAAGTGATT 189
Db 127 TTTCATCATGAAATGATGATGACCTTCTGAGAGATCATTTTCATGAAAAGTGATT 186
QY 190 TGGCTTGGGCGGAATTTTGGAGAAACAGAGATCTGAGTGTATTAAGCGGATTAATAC 249
Db 187 TGGCTTGGGCGGAATTTTGGAGAAACAGAGATCTGAGTGTATTAAGCGGATTAATAC 246
QY 250 ATCCACATGGGTGTTTGGAGTTGGAAGAGTTTGGAAAACAGAGTTTTCACACAG 309
Db 247 ATCCACATGGGTGTTTGGAGTTGGAAGAGTTTGGAAAACAGAGTTTTCACACAG 306
QY 310 CAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 369
Db 307 CAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 366

QY 370 TAAATCAACACGGAACAGAGGTTTCCAAAGAGAGCGGCTATGAGATGGAATTAATTC 429
Db 367 TAAATCAACACGGAACAGAGGTTTCCAAAGAGAGCGGCTATGAGATGGAATTAATTC 426
QY 430 AGAAGCTTCAGGCGCATATACAGAGAGGTTGGAAGAGTAGTTTCCAGGTTCCGTGAGG 489
Db 427 AGAAGCTTCAGGCGCATATACAGAGAGGTTGGAAGAGTAGTTTCCAGGTTCCGTGAGG 486
QY 490 ATTGGTCTAGAAAGTCCAAATTAATGACTTACACCCAGAGCAATGATGACGCGCATCG 549
Db 487 ATTGGTCTAGAAAGTCCAAATTAATGACTTACACCCAGAGCAATGATGACGCGCATCG 546
QY 550 TGGCTTTTGGTCTCTGAGAGACAGATTAATGAGGACAGATTAATGAGTATCTTCTCA 609
Db 547 TGGCTTTTGGTCTCTGAGAGACAGATTAATGAGGACAGATTAATGAGTATCTTCTCA 606
QY 610 AAGCAGAGTGGCAGTGAAGTGAACGAGGTGTTTCAAGAGTTTAAATGAAGATTAAT 669
Db 607 AAGCAGAGTGGCAGTGAAGTGAACGAGGTGTTTCAAGAGTTTAAATGAAGATTAAT 666
QY 670 AACAGGCTCTGGAAGAAATTTCTGGAAGTGAAGAGCAAGAGAGAGAGAAATGATGAT 729
Db 667 AACAGGCTCTGGAAGAAATTTCTGGAAGTGAAGAGCAAGAGAGAGAGAAATGATGAT 726
QY 730 TCAAGGACCAAAAGTGAACCTTACA-TACCCCTCTCTCCACCTGAGATG-AGGATCCCATC 787
Db 727 TCAAGGACCAAAAGTGAACCTTACA-TACCCCTCTCTCCACCTGAGATGAGATCCCATC 786
QY 788 -TTTGACATTATCAGA 803
Db 787 TTTTGACATTATCAGA 803

RESULT 5
CD557866 933 bp mRNA linear EST 11-JUN-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITL
JOURNAL
COMMENT

CD557866 933 bp mRNA linear EST 11-JUN-2003
AGENCOURT_14423254 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30390020 5', mRNA sequence.
CD557866
CD557866.1 GI:31583934
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 933)
NIH-MGC http://imgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINTL at:
http://image.llnl.gov
Plate: NDAM46 row: g column: 21
High quality sequence start: 19
High quality sequence stop: 592.
Location/Qualifiers
1. 933
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390020"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances) "
/clone_lib="NIH_MGC_180"

/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 33.0%; Score 733.8; DB 14; Length 933;
 Best Local Similarity 94.3%; Pred. No. 5.1e-183;
 Matches 816; Conservative 0; Mismatches 42; Indels 7; Gaps 5;

979 ACTGACAGACGAGATTGATGAGCTTGCTCAACAGGCTGCGGAAGACTGCGCTTT 1038
 18 ATTCTCCGGATGAGATTGATGAGCTTGCTCAACAGGCTGCGGAAGACTGCGCTTT 77

1039 TCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1098
 78 TCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 137

1099 GTTGACAGACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
 138 GTTGACAGACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 197

1159 TTGGAAGCCAGAAAATTTCTTTTGGAGCTTGTGTAAGAGCTGTTGTTATATATAGGAG 1218
 198 TTGGAAGCCAGAAAATTTCTTTTGGAGCTTGTGTAAGAGCTGTTGTTATATAGGAG 257

1219 AACCCAGCTGGAGATTCAATTCAGCAAAATGTAAGAGCTGTTGTTATATATAGGAG 1278
 258 AACCCAGCTGGAGATTCAATTCAGCAAAATGTAAGAGCTGTTGTTATATAGGAG 317

1279 TCTGGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338
 318 TCTGGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377

1339 CTTAGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 378 CTTAGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 437

1399 GTTAATTTCTTCCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458
 438 GTTAATTTCTTCCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 497

1459 AACTTTTCCAGAGAAAATTCAGCAAAATGTAAGAGCTGTTGTTATATATAGGAG 1518
 498 AACTTTTCCAGAGAAAATTCAGCAAAATGTAAGAGCTGTTGTTATATATAGGAG 557

1519 TGTGCTGCTGATGAGCAAGTGGTGAAGATGATGATGATGATGATGATGATGATGAT 1578
 558 TGTGCTGCTGATGAGCAAGTGGTGAAGATGATGATGATGATGATGATGATGATGAT 617

1579 TGGCAGAGTTCTCAAAAAGAGAGCTGTTGAATCTGCGAAAATAGGGGATG-AAA 1637
 618 TGGCAGAGTTCTCAAAAAGAGAGCTGTTGAATCTGCGAAAATAGGGGATG-AAA 677

1638 GAACATGATGCTTTTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGAT 1697
 678 GAACATGATGCTTTTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGAT 737

1698 AAGAAAAATATCAACTACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1755
 738 AAGAAAAATATCAACTACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 797

1756 AGCTCTTGG-AGATTTTGGCTTTGAAAAGTGC-CA-GTTCTTGTGTA-CTT-CACTAG 1811
 798 AGCTCTTGGAAATTTTGGCTTTGAAAAGTGC-CA-GTTCTTGTGTA-CTT-CACTAG 857

1812 CTGCCAGAGGGCTGGATATGAAAA 1836
 858 CTGCCAGAGGGCTGGATATGAAAA 882

RESULT 6
 CD557160 885 bp mRNA linear EST 11-JUN-2003
 LOCUS
 DEFINITION
 AGENCOURT_14413590 NIH_MGC_180 Homo sapiens cDNA clone
 IMAGE:30387531 5', mRNA sequence.
 CD557160
 ACCESSION
 CD557160
 VERSION
 CD557160.1 GI:31583228
 KEYWORDS
 EST
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 885)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgsabbs@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at:
 http://image.llnl.gov
 Plate: NDM457 row: p column: 04
 High quality sequence stop: 671.
 Location/Qualifiers
 1..885
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30387531"
 /lab_note="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC_180"
 /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
 Site 2: EcoRV (destroyed); Library is oligo-dt primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.68 kb. Library was
 constructed by (Invitrogen). Note: this is a NIH_MGC
 Library."

FEATURES

source

ORIGIN
 Query Match 32.8%; Score 729.6; DB 14; Length 885;
 Best Local Similarity 99.0%; Pred. No. 6.5e-182;
 Matches 765; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

1223 GACATTCATTCGACAAATAGTACAGGCTGTAATATATATGCTGACTCTGGAAGA 1288
 1 GACATTCATTCGACAAATAGTACAGGCTGTAATATATATGCTGACTCTGGAAGA 60

1289 CTGATGATATCATATGCGCAAGAGAGATGCTCCAAACAGATCAATATCTTATTG 1348
 61 CTGATGATATCATATGCGCAAGAGAGATGCTCCAAACAGATCAATATCTTATTG 120

1349 GATGAAGCTGATGCGATGTTGATATGCTTTGCTCCAGAAATGAAAGTTAATTTCT 1408
 121 GATGAAGCTGATGCGATGTTGATATGCTTTGCTCCAGAAATGAAAGTTAATTTCT 180

1409 TGGCCAGAAATGCCATCAAG 1468
 181 TGGCCAGAAATGCCATCAAG 240

1469 GAGAAATTCAGAGTTGGCTCAGAGTTTAAAGTCAAAATATCTGTTTGTGCTGTT 1528
 241 GAGAAATTCAGAGTTGGCTCAGAGTTTAAAGTCAAAATATCTGTTTGTGCTGTT 300

1529 GAGCAAGTGGGTGAGCATGTAGAGATGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1588

QY 1589 TCAAAAAGAGAAAAAGCTCGTTGAATTCGCGAAACATGAGGGATGAGAAAGTATGATC 1648
Db 361 TCAAAAAGAGAAAAAGCTCGTTGAATTCGCGAAACATGAGGGATGAGAAAGTATGATC 420
QY 1649 TTTGTTAAACTAAGAAAAAGAGATTTTACTGCAACTTTCTTTGTCAAGAAAAATA 1708
Db 421 TTTGTTAAACTAAGAAAAAGAGATTTTACTGCAACTTTCTTTGTCAAGAAAAATA 480
QY 1709 TCAACTCAAGATTCATCGTGTGATCGGAAACAGAGAGCGGAGCAAGCTCTGGAGAT 1768
Db 481 TCAACTCAAGATTCATCGTGTGATCGGAAACAGAGAGCGGAGCAAGCTCTGGAGAT 540
QY 1769 TTTGCTTTGAAAAGTCCCAAGTTCTTTGTTGTTACTTCAAGTCCGAGAGGCTGGAT 1828
Db 541 TTTGCTTTGAAAAGTCCCAAGTTCTTTGTTGTTACTTCAAGTCCGAGAGGCTGGAT 600
QY 1829 ATTGAATATGCAACATGTTATCAATTTTGTATCTTCCATTCATGATGATATGAT 1888
Db 601 ATTGAATATGCAACATGTTATCAATTTTGTATCTTCCATTCATGATGATATGAT 660
QY 1889 CATGGAATGCGGTACTGCTGTGTGGAAATACGAGCAAGCAATTTCC-TTTTGTGA 1947
Db 661 CATGGAATGCGGTACTGCTGTGTGGAAATACGAGCAAGCAATTTCCATTTTGA 720
QY 1948 TCTTGAATGGAT-AAACATTTAGCAAG-CCTCTGTAAGTATTTGACAGA 1998
Db 721 TCTTGAATGGAT-AAACATTTAGCAAG-CCTCTGTAAGTATTTGACAGA 773

RESULT 7
BG717110 798 bp mRNA linear EST 08-MAY-2001
LOCUS 60266919.F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4821503 5',
DEFINITION mRNA sequence.
ACCESSION BG717110
VERSION BG717110.1 GI:13996297
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 798)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs.femail.nih.gov
Tissue Procurement: Miklos Palxovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0728 row: e column: 24
High quality sequence stop: 770.
Location/Qualifiers
1..798
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4821503"
/lab_host="DH10B"
/clone_1b="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
(gagag); Oligo-dT primed using primer
5'-TTTTTTTTTTTAA-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in

ORIGIN
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

Query Match 31.2%; Score 694.4; DB 12; Length 798;
Best Local Similarity 96.3%; Pred. No. 1.3e-172;
Matches 732; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

QY 1 ACTGAAGTCCACATGCGGAGATGAGAAATGCGAAGCAATCAACCTCATATGCTTC 60
Db 39 ACTGAAGTCCACATGCGGAGATGAGAAATGCGAAGCAATCAACCTCATATGCTTC 97
QY 61 CTATGTTCCATATTTGAGAAAGATAGATATTTCTGAGAAATGAGACAAATTTAAAC 120
Db 98 CTATGTTCCATATTTGAGAAAGATAGATATTTCTGAGAAATGAGACAAATTTAAAC 157
QY 121 GATCCAGCTTCATCATGAGAAATGAGATGAGACCTTCGAGAGATCATATTCATGAA 180
Db 158 GATCCAGCTTCATCATGAGAAATGAGATGAGACCTTCGAGAGATCATATTCATGAA 217
QY 181 AAGTGAATTTGCTCTGCGGCGGAAATTTGGAAACAGAGATGCTGATGTAATTAACG 240
Db 218 AAGTGAATTTGCTCTGCGGCGGAAATTTGGAAACAGAGATGCTGATGTAATTAACG 277
QY 241 AGATTAATCATCCCAATGCGTGTGTTTGGAGTTGGAAGATTTTGGAAACAGAGTT 300
Db 278 AGATTAATCATCCCAATGCGTGTGTTTGGAGTTGGAAGATTTTGGAAACAGAGTT 337
QY 301 TTCAACAGCAGGTTTGAAGATGATGATAGCTCTGTTTCTGAGAGAGTCTATATGA 360
Db 338 TTCAACAGCAGGTTTGAAGATGATGATAGCTCTGTTTCTGAGAGAGTCTATATGA 397
QY 361 CTGCGAAGATATATCCACAGGAAACAGAGGTTTTCACAAAGAGCGGCTATGAGATG 420
Db 398 CTGCGAAGATATATCCACAGGAAACAGAGGTTTTCACAAAGAGCGGCTATGAGATG 457
QY 421 AATTAATTCAGAACTTCAGGCGCATACAGAGAGGTGAGAGATGTTCCGAGGTT 480
Db 458 AATTAATTCAGAACTTCAGGCGCATACAGAGAGGTGAGAGATGTTCCGAGGTT 517
QY 481 CCGTGAAGATTTGCTTGAAGAGTCCAAATTAATGACTTAGACCCAGACCAATGATGA 540
Db 518 CCGTGAAGATTTGCTTGAAGAGTCCAAATTAATGACTTAGACCCAGACCAATGATGA 577
QY 541 GCGCAGTGGTGGCTTTTGGTTCTGAGAAACAGATTAAGTGGGACAGGTAAATGTA 600
Db 578 GCGCAGTGGTGGCTTTTGGTTCTGAGAAACAGATTAAGTGGGACAGGTAAATGTA 637
QY 601 TACTTCTCAAGCAGAAAGTGGCAGTGGAGTGAACGAGTGTTCACAAAGGTTAAATGA 660
Db 638 TACTTCTCAAGCAGAAAGTGGCAGTGGAGTGAACGAGTGTTCACAAAGGTTAAATGA 697
QY 661 AGAAGTAATTAACAGGCTCTGGAAGAAATTTCTGAGTCAAGACAGAGAGAGAAAG 720
Db 698 AGAAGTAATTAACAGGCTCTGGAAGAAATTTCTGAGTCAAGACAGAGAGAGAAAG 757
QY 721 T-AGTGAATCTCAAGACCAAAAGTGAAGTCACTATACCC 759
Db 758 T-AGTGAATCTCAAGACCAAAAGTGAAGTCACTATACCC 797

RESULT 8
BI464267 822 bp mRNA linear EST 21-AUG-2001
LOCUS 60320337.F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5269434 5',
DEFINITION mRNA sequence.
ACCESSION BI464267
VERSION BI464267.1 GI:15254923
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 822)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NHL-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovics, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LMML1679 row: m column: 19
High quality sequence stop: 714.
Location/Qualifiers

FEATURES

1. 822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5269434"
/lab_host="DH10B"
/clone_lib="NIH MGC 97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gagag); Oligo-dT primed using primer
5'-TTTTTTTTTTTIVN-3', size-selected for average
insert size 2.2 kb and normalized to R05. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH-MGC Library."

ORIGIN

Query Match 29.3%; Score 652.2; DB 12; Length 822;
Best Local Similarity 95.3%; Pred. No. 2.1e-161;
Matches 737; Conservative 0; Mismatches 28; Indels 8; Gaps 6;

1 ACTTGAAGTACCATGAGGAGTGAAGTGGAGAGCAAGAAATCAACCTCATATGCTTC 60
51 ACTTGAAGTACCATGAGGAGTGAAGTGGAGAGCAAGAAATCAACCTCATATGCTTC 110
61 CTATGTTCCATATTTGAGAGGATAGTATTTCTGAGAGAAATGAGACAAATTTTAAACAG 120
111 CTATGTTCCATATTTGAGAGGATAGTATTTCTGAGAGAAATGAGACAAATTTTAAACAG 170
121 GACTCCAGCTTCATCATGAAATGATGATGACCTTCTGAGAGATCATTTTATGAA 180
171 GACTCCAGCTTCATCATGAAATGATGATGACCTTCTGAGAGATCATTTTATGAA 230
181 AAGTGAATTTGCTCTGGCGGAGATTTTGGAAAGAGATGCTGTATGATTAACGC 240
231 AAGTGAATTTGCTCTGGCGGAGATTTTGGAAAGAGATGCTGTATGATTAACGC 290
241 AGATTAATATCCACAATGAGTGTGTTTGGAGTTGAAAGATTTTGAAGACAGAGTTT 300
291 AGATTAATATCCACAATGAGTGTGTTTGGAGTTGAAAGATTTTGAAGACAGAGTTT 350
301 TTCAAACGACGCTTTGAGAGTGTGATGCTCTGCTTTCTGAGAGAGCTCTGATATGA 360
351 TTCAAACGACGCTTTGAGAGTGTGATGCTCTGCTTTCTGAGAGAGCTCTGATATGA 410
361 CTGGAGAGATATCCACAACGAGACAGAGGCTTTTCCAGAGAGGCGCTATCGAGATG 420
411 CTGGAGAGATATCCACAACGAGACAGAGGCTTTTCCAGAGAGGCGCTATCGAGATG 470
421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGGTGGA-AGAGGTAGTTTCCGAGTT 479
471 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGGTGGA-AGAGGTAGTTTCCGAGTT 530

QY 480 GCCGTGAG-AGGATTTGCTAGGAAGTCCAAATATGACTTAGACCCAGAGATGATG 538
DB 531 GCCGTGAGAGATTTGCTAGGAAGTCCAAATATGACTTAGACCCAGAGATGATG 590
QY 539 CAGCGCACTGAGTGGCTTTTGTCTGAGAGACAGATTAATGAGGACAGATATG 598
DB 591 CAGCGCACTGAGTGGCTTTTGTCTGAGAGACAGATTAATGAGGACAGATATG 649
QY 599 GATCTTTCTCAAGCAGAGTGGCAGTGGAGTGAAGTGAACAGAGTGTATCAAGGTTAAT 658
DB 650 GATCTTTCTCAAGCAGAGTGGCAGTGGAGTGAAGTGAACAGAGTGTATCAAGGTTAAT 709
QY 659 GAAGAGTATATACAGG---CTCTGAAAGATTTCTGAGAGTCAAG-AGGAGAGAGAG 714
DB 710 GAAGAGTATATACAGGCTTTCTGAGAGAGATTTCTGAGAGTCAAGAGAGAGAG 769
QY 715 AGAAGTATGATATCTCAAG-ACCAAGAGTACCTACATACCCCTCTCTCA 766
DB 770 AGAAGTATGATATCCCAAGAGACCAAGAGTACCTACATACCCCTCTCTCA 822

RESULT 9
CD557137 914 bp mRNA linear EST 11-JUN-2003
LOCUS
DEFINITION
AGENCOURT 14413272 NIH MGC 180 Homo sapiens CDNA clone
IMAGE:3030297 5', mRNA sequence.
ACCESSION
CD557137
VERSION
CD557137.1 GI:31583205
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM465 row: c column: 10
High quality sequence start: 42
High quality sequence stop: 582.
Location/Qualifiers

REFERENCE
1 (bases 1 to 914)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NHL-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10X07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM465 row: c column: 10
High quality sequence start: 42
High quality sequence stop: 582.
Location/Qualifiers

FEATURES

1. 914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3030297"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="NIH MGC 180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH-MGC
Library."

ORIGIN

Query Match 29.2%; Score 649; DB 14; Length 914;
Best Local Similarity 96.6%; Pred. No. 1.5e-160;
Matches 706; Conservative 0; Mismatches 20; Indels 5; Gaps 4;

QY 1080 CTGCCAGTGTGTTTAAAGATTGCAAGACCGAGGTATTAATTGTAGCACCACACTCGAG 1139

Db 25 CGATAGTCTTTTAAAGAGTTCAGGAAACCAAGTATATATTTAGACCAACCTCGAG 84

Qy 1140 AATGTGCAACAGATTTA-TTTGGAAGCCAGAAAATTTCTTTGGGACTTGTGTAAGA 1198

Db 85 AATTGGCAACAGATTTATTTTGAAGCCAGAAAATTTCTTTGGGACTTGTGTAAGA 144

Qy 1199 GCTGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCAGCAAAATAGTACAGGC 1258

Db 145 GCTGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCAGCAAAATAGTACAGGC 204

Qy 1259 TGTATATATATATGAGTCTCTCTGGGAAGCTAGTATATCTATAGCAAAAGAAAAGAT 1318

Db 205 TGTATATATATATGAGTCTCTCTGGGAAGCTAGTATATCTATAGCAAAAGAAAAGAT 264

Qy 1319 GGTCTCAACAGATCAAAATCTTATGTTTGGATGAAGCTGATGCAATGTTGATGAGGT 1378

Db 265 GGTCTCAACAGATCAAAATCTTATGTTTGGATGAAGCTGATGCAATGTTGATGAGGT 324

Qy 1379 TTTGGTCCAGAAATGAAGAAATATTTCTTGGCCAGAAATGCAATCAAAAGAAAGAGCC 1438

Db 325 TTTGGTCCAGAAATGAAGAAATATTTCTTGGCCAGAAATGCAATCAAAAGAAAGAGCC 384

Qy 1439 CAACACCTTATGTTCAAGTGCACCTTTCCAGAGAAATTCAAAGTTGCTGCAAGATTT 1498

Db 385 CAACACCTTATGTTCAAGTGCACCTTTCCAGAGAAATTCAAAGTTGCTGCAAGATTT 444

Qy 1499 TTTAAGTCAATATCTGTTTGTGCTTTGTCGACAAAGTGGTGAGCATGTAGATGTT 1558

Db 445 TTTAAGTCAATATCTGTTTGTGCTTTGTCGACAAAGTGGTGAGCATGTAGATGTT 504

Qy 1559 CACGACACGCTTCCAGAGTGGCCAGTTCTCAAAAGAGAAAGAGTCTGTTAAATTCG 1618

Db 505 CACGACACGCTTCCAGAGTGGCCAGTTCTCAAAAGAGAAAGAGTCTGTTAAATTCG 564

Qy 1619 CGAAACATAGGGGATGAAGAACTATGTTGTTGTTGAAGCTAGAAAAGACAGATTTT 1678

Db 565 CGAAACATAGGGGATGAAGAACTATGTTGTTGTTGAAGCTAGAAAAGACAGATTTT 624

Qy 1679 ACTGCAAC-CTTTCTTGTCAAGAAAATATCACTAAGATTCATGATGATC-663 1736

Db 625 ATTGCAACCTTTCTTGTCAAGAAAATATCACTAAGATTCATGATGATGATGATG 684

Qy 1737 AACGAGAGAGCGGAGCAAGCTCTTG--AGATTTTCGCTTTGGAAGTCCGAGTTCT 1794

Db 685 AACGAGAGAGCGGAGCAAGCTCTTGGAATAATTTCCCTTTTGGAAAGTCCGAGTTT 744

Qy 1795 TGTGCTACTT 1805

Db 745 CTGGTTGCT 755

RESULT 10

LOCUS B0429126 792 bp mRNA linear EST 24-MAY-2002

DEFINITION AGENCOURT_7901787 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6104663

ACCESSION B0429126

VERSION B0429126.1 GI:21168202

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 792)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>;

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapds-r@mail.nih.gov

Tissue Procurement: CLONTECH

cDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LICM2342 row: n column: 24

High quality sequence stop: 496.

Location/Qualifiers

1..792

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6104663"

/lab_host="DH10B (TI phage-resistant)"

/clone_id="NIH_MGC_82"

/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccatctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTAATGAC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCAGCATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 28.9%; Score 643.2; DB 13; Length 792;

Best Local Similarity 96.9%; Pred. No. 5e-159;

Matches 677; Conservative 0; Mismatches 19; Indels 3; Gaps 2;

Qy 1 ACTTAAGTCAACATGAGGAGATGAGATGGAAGCAAGAAATCAACCTCATATGCTTC 60

Db 48 ACTTAAGTCAACATGAGGAGATGAGATGGAAGCAAGAAATCAACCTCATATGCTTC 107

Qy 61 CTATGTTCCATATTTGAGAAAGATAGATTTCTGGAAGAAATGAGCAATTTTAAACG 120

Db 108 CTATGTTCCATATTTGAGAAAGATAGATTTCTGGAAGAAATGAGCAATTTTAAACG 167

Qy 121 GATTCAGCTTCATCATCAAGAAATGATGATGACCTTCGAAAGATCATTTCAATGAA 180

Db 168 GATTCAGCTTCATCATCAAGAAATGATGATGACCTTCGAAAGATCATTTCAATGAA 227

Qy 181 AAGTGAATTTCTCTGGGCGGAATTTTGGAAAAGAGATGCTGAGTGAATTAAGCG 240

Db 228 AAGTGAATTTCTCTGGGCGGAATTTTGGAAAAGAGATGCTGAGTGAATTAAGCG 287

Qy 241 AGATTAATCAATCAAGATGAGTGTGTTTGGAGTTGGAAGATTTTGGAAAAGAGATTT 300

Db 288 AGATTAATCAATCAAGATGAGTGTGTTTGGAGTTGGAAGATTTTGGAAAAGAGATTT 347

Qy 301 TTCAACAGAGGTTTGAAGATGATAGTCTGTTTCTGAGAGAGTCTAGTATGA 360

Db 348 TTCAACAGAGGTTTGAAGATGATAGTCTGTTTCTGAGAGAGTCTAGTATGA 407

Qy 361 CTGGAAGATTAATCAACAGGAACAGAGGTTTCCAGAGAGGCGCTATGAGATGG 420

Db 408 CTGGAAGATTAATCAACAGGAACAGAGGTTTCCAGAGAGGCGCTATGAGATGG 467

Qy 421 AAATTAATTCAGAGCTTCAGAGGCGCATACAGAAAGAGTGAAGAGTATTCGAGGTTG 480

Db 468 AAATTAATTCAGAGCTTCAGAGGCGCATACAGAAAGAGTGAAGAGTATTCGAGGTTG 527

Qy 481 CCGTGAAGATTTGCTCTAGAGAGTCCAAATTAATGATTAAGACCCAGAGAAATGATGA 540

Db 528 CCGTGAAGATTTGCTCTAGAGAGTCCAAATTAATGATTAAGACCCAGAGAAATGATGA 587

Qy 541 GCGCAGCTGGG--CGTTTGTGTTTGAAGAACCAATTAATGAGTGAAGTGAAGGTTG 599

Db 588 GCGCAGCTGGGCGCTTTTGTGTTTGAAGAACCCATTAATGAGTGAAGTGAAGGTTG 647

Qy 600 ATACTTCTCAAGCAGAGTGGCAGTGAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGA 659

Db 648 ATACTTCTCAAGCAGAGTGGCAGTGAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGA 707

QY 660 AAGAGTAATACAGG--CTCTGGAAAAGATTCTTGGAA 696
 Db 708 AAGAGTAATACAGGCGCTCTGGAAAAGATTCTTGGAA 746

RESULT 11
 BG719449 747 bp mRNA linear EST 08-MAY-2001
 LOCUS 602690172F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822528 5',
 DEFINITION mRNA sequence.
 BG719449
 VERSION BG719449.1 GI:13998636
 SOURCE EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 cDNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.jnl.gov
 Plate: LHAM10730 row: p column: 17
 High quality sequence stop: 734.
 Location/Qualifiers

FEATURES
 source
 1..747
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4822528"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
 (gtcgag); Oligo-dT primed using primer
 5'-TTTTTTTTTTTTTTN-3', size-selected for average
 insert size 2.2 kb and normalized to ROI 5. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein
 (NIH/NHGRI, National Institutes of Health). Note: this is
 a NIH_MGC Library."

ORIGIN
 Query Match 28.6%; Score 636.2; DB 12; Length 747;
 Best Local Similarity 99.2%; Pred. No. 3.5e-157;
 Matches 660; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 ACTGAGTCCATGGGGGATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTTC 60
 Db 84 ACTGAAAGCCACATGGAGATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTTC 143
 QY 61 CTATGTTCCATATTTGGAAGATAGTATTTGGAAGAAATGAGACAAATTTTAAACAG 120
 Db 144 CTATGTTCCATATTTGGAAGATAGTATTTGGAAGAAATGAGACAAATTTTAAACAG 203
 QY 121 GACTCCAGCTTCATCAGAAATGAGATGAGACCTTTCCGAGAGATCATTTTCAAGAA 180
 Db 204 GACTCCAGCTTCATCAGAAATGAGATGAGACCTTTCCGAGAGATCATTTTCAAGAA 263
 QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGTGAGTGAATAAGCG 240
 Db 264 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGTGAGTGAATAAGCG 323

QY 241 AGATATATCATCCACATGSGTGTGTTTGGAGTTGGAAAAGTTTGGAAAACAGAGTTT 300
 Db 324 AGATATATCATCCACATGSGTGTGTTTGGAGTTGGAAAAGTTTGGAAAACAGAGTTT 383
 QY 301 TTCAAACGACAGGTTTGAAGATGAGTAGCTGTGTTTCTGGAGAGAGTCAATGATGA 360
 Db 384 TTCAAACGACAGGTTTGAAGATGAGTAGCTGTGTTTCTGGAGAGAGTCAATGATGA 443
 QY 361 CTGCGAAGATATATCCACACGGAACAGAGGTTTCCAAAGAGCGGCTTATCGAGATGG 420
 Db 444 CTGCGAAGATATATCCACACGGAACAGAGGTTTCCAAAGAGCGGCTTATCGAGATGG 503
 QY 421 AA-ATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGGTAGTATTCGAGGTT 479
 Db 504 AACATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGGTAGTATTCGAGGTT 563
 QY 480 GCCGTGAGAGATTGGTCTAGAGAGTCCAAATTAATGACTTAGACCACAGCATGATATGC 539
 Db 564 GCCGTGAGAGATTGGTCTAGAGAGTCCAAATTAATGACTTAGACCACAGCATGATATGC 623
 QY 540 AGCGCACTGTGGCTTTTGGTTCTTGAAGACAGATTTAAGTGCACAGTAAATGTTG 599
 Db 624 AGCGCACTGTGGCTTTTGGTTCTTGAAGACAGATTTAAGTGCACAGTAAATGTTG 683
 QY 600 ATACTTCTCAAAGCAGAAAGTGCAGTGAAGTGAACGAGGTGTTACAAAGTTTAAATG 659
 Db 684 ATACTTCTCAAAGCAGAAAGTGCAGTGAAGTGAACGAGGTGTTACAAAGTTTAAATG 742

QY 660 AAGAA 664
 Db 743 AAGAA 747

RESULT 12
 BG718234 682 bp mRNA linear EST 08-MAY-2001
 LOCUS 60269209F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828298 5',
 DEFINITION mRNA sequence.
 BG718234
 VERSION BG718234.1 GI:13997421
 SOURCE EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 cDNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.jnl.gov
 Plate: LHAM10746 row: a column: 03
 High quality sequence stop: 682.
 Location/Qualifiers

FEATURES
 source
 1..682
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4828298"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
 (gtcgag); Oligo-dT primed using primer

ORIGIN

5'-TTTTTTTTTTTTTTN-3', size-selected for average insert size 2.2 kb and normalized to R05. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHRI, National Institutes of Health). Note: this is a NIH_MGC Library."

Query Match 28.4%; Score 631.8; DB 12; Length 682;
Best Local Similarity 99.7%; Pred. No. 4.9e-156;
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACTTGAAGTACCATGGGGATGAAATGGGAGGAGAAATCAACCTCATATGCTTC 60
DB 48 ACTTGAAGTACCATGGGGATGAAATGGGAGGAGAAATCAACCTCATATGCTTC 107
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATTCGAGAAATGAGACATTTTAACAG 120
DB 108 CTATGTTCCCATATTTGAGAGAGATAGTATTCGAGAAATGAGACATTTTAACAG 167
QY 121 GATCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 180
DB 168 GATCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 227
QY 181 AAGTGAATTCCTCTGGGCGGAAATTTGGAACAGAGATGCTGTGATGATTAAGCG 240
DB 228 AAGTGAATTCCTCTGGGCGGAAATTTGGAACAGAGATGCTGTGATGATTAAGCG 287
QY 241 AGATTAATCATCAACATGGGTTGTTGAGTTGGAAGAGTTTGGAAACAGAGCTTT 300
DB 288 AGATTAATCATCAACATGGGTTGTTGAGTTGGAAGAGTTTGGAAACAGAGCTTT 347
QY 301 TTCAAAACAGAGTTTGAAGATGATGATGCTGTGTTCTGGAGAGATCATTAATGA 360
DB 348 TTCAAAACAGAGTTTGAAGATGATGATGCTGTGTTCTGGAGAGATCATTAATGA 407
QY 361 CTCGGAAGATTAATCAACAGAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 420
DB 408 CTCGGAAGATTAATCAACAGAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 467
QY 421 AATTAATTCAGAGCTTCAGGCGCATCAAGAGAGTGGAAAGATGTTCCGAGTTG 480
DB 468 AATTAATTCAGAGCTTCAGGCGCATCAAGAGAGTGGAAAGATGTTCCGAGTTG 527
QY 481 CCGTGAAGATTTGCTCTAGAGATCAATTAATGACTTACCCAGACGATGATGA 540
DB 528 CCGTGAAGATTTGCTCTAGAGATCAATTAATGACTTACCCAGACGATGATGA 587
QY 541 GCGCACTGGTGGCTTTTGGTTCTAGAGACCAATTAATGAGTGAACAGTATGGA 600
DB 588 GCGCACTGGTGGCTTTTGGTTCTAGAGACCAATTAATGAGTGAACAGTATGGA 647
QY 601 TACTTCTCAAGCAGAGTGGCAGTGAAGTGAAC 635
DB 648 TACTTCTCAAGCAGAGTGGCAGTGAAGTGAAC 682

RESULT 13
BIS60849
LOCUS 603254018F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296444 5',
DEFINITION
tRNA sequence.
ACCESSION BIS60849
VERSION BIS60849.1 GI:15448163
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 693)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Mixlos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LHAM1750 row: c column: 05
High quality sequence stop: 691.

FEATURES

source

Location/Qualifiers
1..693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5296444"
/lab_host="DH10B"
/clone_lib="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
(gtagag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTN-3', size-selected for average
insert size 2.2 kb and normalized to R05. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN

Query Match 28.4%; Score 631.8; DB 12; Length 693;
Best Local Similarity 99.7%; Pred. No. 5e-156;
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACTTGAAGTACCATGGGGATGAAATGGGAGGAGAAATCAACCTCATATGCTTC 60
DB 59 ACTTGAAGTACCATGGGGATGAAATGGGAGGAGAAATCAACCTCATATGCTTC 118
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATTCGAGAAATGAGACATTTTAACAG 120
DB 119 CTATGTTCCCATATTTGAGAGAGATAGTATTCGAGAAATGAGACATTTTAACAG 178
QY 121 GATCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 180
DB 179 GATCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 238
QY 181 AAGTGAATTCCTCTGGGCGGAAATTTGGAACAGAGATGCTGTGATGATTAAGCG 240
DB 239 AAGTGAATTCCTCTGGGCGGAAATTTGGAACAGAGATGCTGTGATGATTAAGCG 298
QY 241 AGATTAATCATCAACATGGGTTGTTGAGTTGGAAGAGTTTGGAAACAGAGCTTT 300
DB 299 AGATTAATCATCAACATGGGTTGTTGAGTTGGAAGAGTTTGGAAACAGAGCTTT 358
QY 301 TTCAAAACAGAGTTTGAAGATGATGATGCTGTGTTCTGGAGAGATCATTAATGA 360
DB 359 TTCAAAACAGAGTTTGAAGATGATGATGCTGTGTTCTGGAGAGATCATTAATGA 418
QY 361 CTCGGAAGATTAATCAACAGAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 420
DB 419 CTCGGAAGATTAATCAACAGAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 478
QY 421 AATTAATTCAGAGCTTCAGGCGCATCAAGAGAGTGGAAAGATGTTCCGAGTTG 480
DB 479 AATTAATTCAGAGCTTCAGGCGCATCAAGAGAGTGGAAAGATGTTCCGAGTTG 538
QY 481 CCGTGAAGATTTGCTCTAGAGACCAATTAATGACTTACCCAGACGATGATGA 540
DB 539 CCGTGAAGATTTGCTCTAGAGACCAATTAATGACTTACCCAGACGATGATGA 598

QY 541 GCGACACTGGGCTTTTGGTTCTAGAACACGATTAAGTGGACAGATATGGTGA 600
 DB 599 GCGACACTGGGCTTTTGGTTCTAGAACACGATTAAGTGGACAGATATGGTGA 658
 QY 601 TACTTCTCAAGCAGAGAGTGGCAGTGGAGTGAAC 635
 DB 659 TACTTCTCAAGCAGAGAGTGGCAGTGGAGTGAAC 693

RESULT 14
 CD358073 868 bp mRNA linear EST 29-MAY-2003
 LOCUS AGENCOURT 14257996 NIH_MGC.180 Homo sapiens CDNA clone
 DEFINITION IMAGE:30386875 5', mRNA sequence.
 ACCESSION CD358073 GI:31129484
 VERSION CD358073.1 GI:31129484
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 868)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 CDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.lnl.gov
 Plate: NDAM456 row: d column: 20
 High quality sequence stop: 578.
 Location/Qualifiers
 1..868
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30386875"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH MGC 180"
 /note="Organ: Testis; Vector: pCW-SPORT6.1; Site 1: NotI;
 Site 2: EcoRV (destroyed); Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.68 Kb. Library was
 constructed by (Invitrogen). Note: this is a NIH_MGC
 Library."

ORIGIN
 Query Match 28.4%; Score 631.2; DB 14; Length 868;
 Best Local Similarity 96.6%; Pred. No. 7.9e-156;
 Matches 677; Conservative 0; Mismatches 19; Indels 5; Gaps 3;

QY 1 ACTGAAGTCAACATGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 60
 DB 45 ACTGAAGTCAACATGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 104
 QY 61 CTATGTTCCATATTGAGAGATAGTATTCTGAGAAATGAGCAATTTTAACAG 120
 DB 105 CTATGTTCCATATTGAGAGATAGTATTCTGAGAAATGAGCAATTTTAACAG 164
 QY 121 GACTCAGCTTCATCATGAGAAATGATGAGACCTTCGAGAGATCATTTTCATGAA 180
 DB 165 GACTCAGCTTCATCATGAGAAATGATGAGACCTTCGAGAGATCATTTTCATGAA 224
 QY 181 AAGTGATTTGCTCTGGCGGAAATTTGAGAAAGAGATGCTGATGTAATAGCG 240

DB 225 AAGTGATTTGCTCTGGCGGAAATTTGAGAAAGAGATGCTGATGTAATAGCG 284
 QY 241 AATATATCATTCACAAATGGGTTGTTTGGAGTTGGAAAGAGTTTGGAAACAGAGTTT 300
 DB 285 AATATATCATTCACAAATGGGTTGTTTGGAGTTGGAAAGAGTTTGGAAACAGAGTTT 344
 QY 301 TTCAAAGCAGAGTTTGAAGATGATAGTCTGCTGTTTTCGAGAGATCTAGTAATGA 360
 DB 345 TTCAAAGCAGAGTTTGAAGATGATAGTCTGCTGTTTTCGAGAGATCTAGTAATGA 404
 QY 361 CTGCGAAGATTAATCCACACGAGACAGAGGTTTTCACAGAGAGCGGCTATCGAGATGG 420
 DB 405 CTGCGAAGATTAATCCACACGAGACAGAGGTTTTCACAGAGAGCGGCTATCGAGATGG 464
 QY 421 AATATATCAGAGTTTCAAGGTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 DB 465 AATATATCAGAGTTTCAAGGTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
 QY 481 CCGTGGAGATTTGGTCTAGAGAGTCCAAATATGACTTAGACCCAGACGAATGTATGCA 540
 DB 525 CCGTGGAGATTTGGTCTAGAGAGTCCAAATATGACTTAGACCCAGACGAATGTATGCA 584
 QY 541 GCGCACTGGTGGCTTTTGGTTCTAGAGAGACCAATATTAAGTGGCAGAGTAATAGTGA 600
 DB 585 GCGCACTGGTGGCTTTTGGTTCTAGAGAGACCAATATTAAGTGGCAGAGTAATAGTGA 644
 QY 601 TACTTCTCAAGCAGAGATGG-CAGTGGAGTGAACAGAGTGTATC--AAAGTTTAA 657
 DB 645 TACTTCTCAAGCAGAGATGGAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
 QY 658 TGAAGAGTAATPAACG--CTCTGAGAAAGAAATTTTGA 696
 DB 705 GGAAGAGTAATPAACGCGGCTCTGAGAAAGAAATTTTGA 745

RESULT 15
 B1462207 830 bp mRNA linear EST 21-AUG-2001
 LOCUS 60305520F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:5271221 5',
 DEFINITION mRNA sequence.
 ACCESSION B1462207
 VERSION B1462207.1 GI:15252863
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 830)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.lnl.gov
 Plate: LLM11684 row: h column: 06
 High quality sequence stop: 769.
 Location/Qualifiers
 1..830
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5271221"
 /lab_host="DH10B"
 /clone_lib="NIH MGC_97"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI

(gscgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.2 kb and normalized to ROP 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN

```
Query Match      28.2%; Score 627; DB 12; Length 830;
Best Local Similarity 95.7%; Pred. No. 1e-154;
Matches 741; Conservative 0; Mismatches 21; Indels 12; Gaps 9;

QY 1 ACTTGAAGTACCATGGGGATGAGATTGGGAAGCAGAAATCAACCTCATATGCTTC 60
   |||
Db 59 ACTTGAAGCCACCATGGAGATGAGATTGGGAAGCAGAAATCAACCTCATATGCTTC 118
   |||

QY 61 CTATGTTCCCATATTATTGAGAAGATAGTATTCTGAGAAAAATGAGACAAATTTAACG 120
   |||
Db 119 CTATGTTCCCATATTATTGAGAAGATAGTATTCTGAGAAAAATGAGACAAATTTAACG 178
   |||

QY 121 GACTCCAGCTTCATCATCAGAAATGATGAGACCTCTCGAAGAGATCATTTATGAA 180
   |||
Db 179 GACTCCAGCTTCATCATCAGAAATGATGAGACCTCTCGAAGAGATCATTTATGAA 238
   |||

QY 181 AAGTGATTTGCTCTGGGCGGAATTTGGAAAAGAGATGCTGTGAGTATATAAGCG 240
   |||
Db 239 AAGTGATTTGCTCTGGGCGGAATTTGGAAAAGAGATGCTGTGAGTATATAAGCG 298
   |||

QY 241 AGATTAATACATCCAAATGGGTGTTTGGAGTTGGAAGAAGTTTGGAAACAGAGTTT 300
   |||
Db 299 AGATTAATACATCCAAATGGGTGTTTGGAGTTGGAAGAAGTTTGGAAACAGAGTTT 358
   |||

QY 301 TTCAACAGCAGGTTTGAGATGATGATAGCTGTGTTTCTGAGAGAGCTAGTATAGA 360
   |||
Db 359 TTCAACAGCAGGTTTGAGATGATGATAGCTGTGTTTCTGAGAGAGCTAGTATAGA 418
   |||

QY 361 CTGGGAAGATTAATCCAAACAGGAAGAGGTTTCCAAAGAGGCGGCTATCGAGATGG 420
   |||
Db 419 CTGGGAAGATTAATCCAAACAGGAAGAGGTTTCCAAAGAGGCGGCTATCGAGATGG 478
   |||

QY 421 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTAGTTCCGAGGTTG 480
   |||
Db 479 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTAGTTCCGAGGTTG 538
   |||

QY 481 CCGTGGAGGATTTGGTCTAGGAAGTCGAATTAATGACTTAGACCCAGACGAATGTATGCA 540
   |||
Db 539 CCGTGGAGGATTTGGTCTAGGAAGTCGAATTAATGACTTAGACCCAGACGAATGTATGCA 597
   |||

QY 541 GCGCAGCTGGTGCTTTTGGTCTTAGAAGACCAAGTATTAGT-GGCACAGGTAATGGTG 599
   |||
Db 598 GCGCAGCTGGTGCTTTTGGTCTTAGAAGACCAAGTATTAGTGGGACAGGTAATGGTG 656
   |||

QY 600 ATAATTCTCAAGCAGAAAGTGGCAG-TGGAAGTGAACG-AGGTGTTACAAAGGTTT-- 654
   |||
Db 657 ATAATTCTCAAGCAGAAAGTGGCAGTGGGAAGTGAACGAAGGTGTTACAAAGGTTTCA 716
   |||

QY 655 AAATGAAGAGTATATAACAGGCTCTGGAA-AGAATCTTGGAAAGTCAGAA-GCAGAAGGA 712
   |||
Db 717 ACTGAACGAAGTATATAACAGGCTCTGGAA-AGAATCTTGGAAAGTCAGAAAGCAGAAGGA 776
   |||

QY 713 GGA-GAAAGTACTGATACTCAAGAGACCAAAAGTGACTTACATACCCCTCTCTC 764
   |||
Db 777 GGAAGAAAGTACTGATACTCAAGAGACCAAAAGTGAACTTAATACCCCTCTCTC 830
   |||
```

Search completed: July 27, 2004, 10:49:42
Job time : 5709 secs